



# **STIC Search Report**

## **Biotech-Chem Library**

STIC Database Tracking Number: 116350

TO: Sheridan Swope  
Location: REM-3C70/3A70  
Art Unit: 1652  
Monday, March 15, 2004

Case Serial Number: 09/966880

From: Barb O'Bryen  
Location: Biotech-Chem Library  
Remsen E01A69  
Phone: 571-272-2518

*BOB*  
barbara.obryen@uspto.gov

### Search Notes

# RUSH



		<u>Needed</u>	<u>Done</u>	<u>IS</u>
SID 7	comprising	FL	✓	Need ✓
80-676	consisting oligo	oligo	✓	Need ✓
SID 8	comprising	FL	✓	Need ✓
	consisting oligo	oligo	✓	✓
SID 11-15	consisting oligo	FL oligo No oligo on SID 15 (poly A)	✓ ✓	12-14 Need FL encompassed in oligo SID 7 search
SID 9	consisting U	FL D... U	✓ -	Need ✓ "
SID 10	consisting n D	FL	Need ✓	Need ✓
SID 35	consisting comprising	FL	✓	✓



Seq 15 is 3741 - 5912 of Seq 10  
Seq 13 is 2592 - 2862 of Seq 10  
Seq 12 is 1065 - 1212 of Seq 10  
Seq 14 is 3156 - 3271 of Seq 10

Seq 11 is 1032 - 1118 of Seq 9 =

SH 61.30  
Seq 9 - interference - 10 hrs  
Seq 10, 11 - All NA regular - 11 hrs  
" " oligomer 11 hrs

65 hours

2 day



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 12, 2004, 12:22:03 ; Search time 194 Seconds

(without alignments)  
11329.453 Million cell updates/sec

Title: US-09-966-880A-7\_COPY\_80\_676

Perfect score: 597

Sequence: 1 atggacagcctcttgatgaa.....cttcactcttgagacttga 597

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 243257 seqs, 184079884 residues

Word size : 20

Total number of hits satisfying chosen parameters: 27

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :

Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_NEW\_PUB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 17: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 18: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARYS

Result No.	Score	Query Match Length	ID	Description
1	597	100.0	2818	US-09-966-880A-7 Sequence 7, Appl
2	273	45.7	6564	US-09-966-880A-10 Sequence 10, Appl
3	273	45.7	11204	US-09-966-880A-35 Sequence 35, Appl
4	271	45.4	271	US-09-966-880A-13 Sequence 13, Appl
5	148	24.8	148	US-09-966-880A-12 Sequence 12, Appl
6	78	13.1	116	US-09-966-880A-14 Sequence 14, Appl
7	54	9.0	2172	US-09-966-880A-15 Sequence 15, Appl
8	32	5.4	2440	US-09-966-880A-1 Sequence 1, Appl
9	30	5.0	30	US-09-966-880A-18 Sequence 18, Appl
10	30	5.0	30	US-09-966-880A-24 Sequence 24, Appl
11	28	4.7	28	US-09-966-880A-25 Sequence 25, Appl
12	28	4.7	28	US-09-966-880A-28 Sequence 28, Appl
13	27	4.5	27	US-09-966-880A-19 Sequence 19, Appl
14	371	10	371	US-09-918-995-30237 Sequence 30237, A
15	22	3.7	553	US-09-796-692-4504 Sequence 4504, Ap

16	22	3.7	553	14	US-10-040-862-4504	Sequence 4504, Ap
17	22	3.7	553	15	US-10-057-475B-4504	Sequence 4504, Ap
18	22	3.7	553	15	US-10-154-884B-4504	Sequence 4504, Ap
19	22	3.7	556	9	US-09-796-692-5061	Sequence 5061, Ap
20	22	3.7	556	14	US-10-040-862-5061	Sequence 5061, Ap
21	22	3.7	556	15	US-10-057-475B-5061	Sequence 5061, Ap
22	22	3.7	556	15	US-10-154-884B-5061	Sequence 5061, Ap
23	22	3.7	1155	15	US-10-460-923-1	Sequence 1, Appl
24	22	3.7	1534	9	US-09-728-674-173	Sequence 173, Appl
25	22	3.7	2151	9	US-09-822-830A-359	Sequence 359, Appl
26	21	3.5	21	9	US-09-966-880A-26	Sequence 26, Appl
27	20	3.4	63158	14	US-10-292-198-1	Sequence 1, Appl

#### ALIGNMENTS

RESULT 1  
US-09-966-880A-7  
Sequence 7, Application US/09966880A  
Patent No. US20020164743A1  
GENERAL INFORMATION:  
APPLICANT: Honjo, Tasuku  
APPLICANT: Muramatsu, Masamichi  
TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
FILE REFERENCE: 06501-088001  
CURRENT APPLICATION NUMBER: US/09/966, 880A  
PRIOR FILING DATE: 2001-09-28  
PRIOR APPLICATION NUMBER: PCT/JP00/01918  
PRIOR FILING DATE: 2000-03-28  
PRIOR APPLICATION NUMBER: JP 11-371382  
PRIOR FILING DATE: 1999-12-27  
PRIOR APPLICATION NUMBER: JP 11-178999  
PRIOR FILING DATE: 1999-06-24  
PRIOR APPLICATION NUMBER: JP 11-87192  
PRIOR FILING DATE: 1999-03-29  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 7  
LENGTH: 2818  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (80)...(673)  
FEATURE:  
NAME/KEY: 5'UTR  
LOCATION: (1)...(79)  
FEATURE:  
NAME/KEY: 3'UTR  
LOCATION: (677)...(2818)  
US-09-966-880A-7

Query Match 100.0%; Score 597; DB 9; Length 2818;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATGACAGCCTCTTGATGAAACCGAGAGAGTTTCTTACCAATTCAAAATGTCGGCTGG	60
DB	80	ATGACAGCCTCTTGATGAAACCGAGAGAGTTTCTTACCAATTCAAAATGTCGGCTGG	139
QY	61	GCTAAGGTCGGCTGAGACCTTACTGTGCTAGTGAAGAGCGCTGACAGTCTACA	120
DB	140	GCTAAGGTCGGCTGAGACCTTACTGTGCTAGTGAAGAGCGCTGACAGTCTACA	199
QY	121	TCCTTTCACTGACCTTGGTTATCTTGCAGATTAAGACGGCTGCCAGTGAATTGCTC	180
DB	200	TCCTTTCACTGACCTTGGTTATCTTGCAGATTAAGACGGCTGCCAGTGAATTGCTC	259
QY	181	TTCTCTCGCTACATCTGGAAGTGGACCTTGAAGCTGCTGCTACCGCGTCACTGG	240
DB	260	TTCTCTCGCTACATCTGGAAGTGGACCTTGAAGCTGCTGCTACCGCGTCACTGG	319



Query Match 45.7%; Score 273; DB 9; Length 6564;  
Best Local Similarity 100.0%; Pred. No. 2.7e-139;  
Matches 273; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query 155 AGAAGGCTCCACGAGGATTTGCTCTTCTCCGCTACATCTTCGAGCTGAGACCTAGACC 214  
Db 2590 AGAAGGCTCCACGAGGATTTGCTCTTCTCCGCTACATCTTCGAGCTGAGACCTAGACC 2649

Query 215 CTGAGCCGCTCTACCGGCTCACTGTGTTCACTCTGAGAGCCCTGCTAGACTGAGTGGCC 274  
Db 2650 CTGAGCCGCTCTACCGGCTCACTGTGTTCACTCTGAGAGCCCTGCTAGACTGAGTGGCC 2709

Query 275 GACATGTGGCGGACTTTCTGCGAGGGAACCCCACTCACTGAGGATCTTACCGCGC 334  
Db 2710 GACATGTGGCGGACTTTCTGCGAGGGAACCCCACTCACTGAGGATCTTACCGCGC 2769

Query 335 GCCTTACTTCTGTGAGGACCGCAAGGCTAGGCCCGAGGGGCTGGCGGCTGCAACGGC 394  
Db 2770 GCCTTACTTCTGTGAGGACCGCAAGGCTAGGCCCGAGGGGCTGGCGGCTGCAACGGC 2829

Query 395 CCGGGGTGCAATATGACCATATGACCTTAAAG 427

RESULT 2  
US-09-966-880A-10  
Sequence 10, Application US/09966880A  
Patent No. US20020164743A1  
GENERAL INFORMATION:  
APPLICANT: Honjo, Tasuku  
APPLICANT: Muramatsu, Masamichi  
TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
FILE REFERENCE: 06501-088001  
CURRENT APPLICATION NUMBER: US/09/966,880A  
CURRENT FILING DATE: 2001-09-28  
PRIOR APPLICATION NUMBER: PCT/JP00/01918  
PRIOR FILING DATE: 2000-03-28  
PRIOR APPLICATION NUMBER: JP 11-371382  
PRIOR FILING DATE: 1999-12-27  
PRIOR APPLICATION NUMBER: JP 11-178999  
PRIOR FILING DATE: 1999-06-24  
PRIOR APPLICATION NUMBER: JP 11-87192  
PRIOR FILING DATE: 1999-03-29  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 10  
LENGTH: 6564  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-966-880A-10

Query Match 45.7%; Score 273; DB 9; Length 11204;  
Best Local Similarity 100.0%; Pred. No. 2.8e-139;  
Matches 273; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query 155 AGAAGGCTCCACGAGGATTTGCTCTTCTCCGCTACATCTTCGAGCTGAGACCTAGACC 214  
Db 7805 AGAAGGCTCCACGAGGATTTGCTCTTCTCCGCTACATCTTCGAGCTGAGACCTAGACC 7864

Query 215 CTGAGCCGCTCTACCGGCTCACTGTGTTCACTCTGAGAGCCCTGCTAGACTGAGTGGCC 274  
Db 7865 CTGAGCCGCTCTACCGGCTCACTGTGTTCACTCTGAGAGCCCTGCTAGACTGAGTGGCC 7924

Query 275 GACATGTGGCGGACTTTCTGCGAGGGAACCCCACTCACTGAGGATCTTACCGCGC 334  
Db 7925 GACATGTGGCGGACTTTCTGCGAGGGAACCCCACTCACTGAGGATCTTACCGCGC 7984

Query 335 GCCTTACTTCTGTGAGGACCGCAAGGCTAGGCCCGAGGGGCTGGCGGCTGCAACGGC 394  
Db 7985 GCCTTACTTCTGTGAGGACCGCAAGGCTAGGCCCGAGGGGCTGGCGGCTGCAACGGC 8044

Query 395 CCGGGGTGCAATATGACCATATGACCTTAAAG 427  
Db 8045 CCGGGGTGCAATATGACCATATGACCTTAAAG 8077

RESULT 3  
US-09-966-880A-13  
Sequence 13, Application US/09966880A  
Patent No. US20020164743A1  
GENERAL INFORMATION:  
APPLICANT: Honjo, Tasuku  
APPLICANT: Muramatsu, Masamichi  
TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
FILE REFERENCE: 06501-088001  
CURRENT APPLICATION NUMBER: US/09/966,880A  
CURRENT FILING DATE: 2001-09-28  
PRIOR APPLICATION NUMBER: PCT/JP00/01918  
PRIOR FILING DATE: 2000-03-28  
PRIOR APPLICATION NUMBER: JP 11-371382  
PRIOR FILING DATE: 1999-12-27  
PRIOR APPLICATION NUMBER: JP 11-178999  
PRIOR FILING DATE: 1999-06-24



;; PRIOR APPLICATION NUMBER: JP 11-87192  
;; PRIOR FILING DATE: 1999-03-29  
;; NUMBER OF SEQ ID NOS: 36  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO: 13  
;; LENGTH: 271  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
US-09-966-880A-13

Query Match 45.4%; Score 271; DB 9; Length 271;  
Best Local Similarity 100.0%; Pred. No. 3.2e-158;  
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 157 AACGGCTGCACGAGTAATGCTCTTCCGCTACATCTGGAGCTGGAGACCTGACCTT 216  
DB 1 AACGGCTGCACGAGTAATGCTCTTCCGCTACATCTGGAGCTGGAGACCTGACCTT 60  
QY 217 GGCCGCTGCTACCGGCTGACCTGTTCACTCTCTGAGACCCCTGCTACGACTGACCGCA 276  
DB 61 GGCCGCTGCTACCGGCTGACCTGTTCACTCTCTGAGACCCCTGCTACGACTGACCGCA 120  
QY 277 CATGTCGCGCACTTCTCGAGAGGAACCCCAACTCACTGAGAGATCTTACCGCGCGC 336  
DB 121 CATGTCGCGCACTTCTCGAGAGGAACCCCAACTCACTGAGAGATCTTACCGCGCGC 180  
QY 337 CTCTACTTCTGTGAGAGCCGCAAGGCTGAGCCGAGGGGCTGGCGGCTGACCGCGCC 396  
DB 181 CTCTACTTCTGTGAGAGCCGCAAGGCTGAGCCGAGGGGCTGGCGGCTGACCGCGCC 240  
QY 397 GGGGTGCAAAATGACATCATGACCTTCAAG 427  
DB 241 GGGGTGCAAAATGACATCATGACCTTCAAG 271

RESULT 5  
US-09-966-880A-12

;; Sequence 12, Application US/09966880A  
;; Patent No. US20020164743A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Honjo, Tasuku  
;; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
;; FILE REFERENCE: 06501-088001  
;; CURRENT APPLICATION NUMBER: US/09/966, 880A  
;; PRIOR FILING DATE: 2001-09-28  
;; PRIOR APPLICATION NUMBER: PCT/JP00/01918  
;; PRIOR FILING DATE: 2000-03-28  
;; PRIOR APPLICATION NUMBER: JP 11-371382  
;; PRIOR FILING DATE: 1999-12-27  
;; PRIOR APPLICATION NUMBER: JP 11-178999  
;; PRIOR FILING DATE: 1999-06-24  
;; PRIOR APPLICATION NUMBER: JP 11-87192  
;; PRIOR FILING DATE: 1999-03-29  
;; NUMBER OF SEQ ID NOS: 36  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO: 12  
;; LENGTH: 148  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
US-09-966-880A-12

Query Match 24.8%; Score 148; DB 9; Length 148;  
Best Local Similarity 100.0%; Pred. No. 1.2e-70;  
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 CCTCTGATGAACCGAGAAATTTCTTACCAATCAAAAATGTCCGTCGCGTAAGG 68  
DB 1 CCTCTGATGAACCGAGAAATTTCTTACCAATCAAAAATGTCCGTCGCGTAAGG 60  
QY 69 TCGGCGTGAAGACTTACCTGTGCTAGTGAAGAGGCGTGAACAATGCTTCTTTTC 128  
DB 61 TCGGCGTGAAGACTTACCTGTGCTAGTGAAGAGGCGTGAACAATGCTTCTTTTC 120

QY 129 ACTGACCTTGTGTTATCTTGGCAATAG 156  
DB 121 ACTGACCTTGTGTTATCTTGGCAATAG 148

RESULT 6  
US-09-966-880A-14

;; Sequence 14, Application US/09966880A  
;; Patent No. US20020164743A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Honjo, Tasuku  
;; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
;; FILE REFERENCE: 06501-088001  
;; CURRENT APPLICATION NUMBER: US/09/966, 880A  
;; PRIOR FILING DATE: 2001-09-28  
;; PRIOR APPLICATION NUMBER: PCT/JP00/01918  
;; PRIOR FILING DATE: 2000-03-28  
;; PRIOR APPLICATION NUMBER: JP 11-371382  
;; PRIOR FILING DATE: 1999-12-27  
;; PRIOR APPLICATION NUMBER: JP 11-178999  
;; PRIOR FILING DATE: 1999-06-24  
;; PRIOR APPLICATION NUMBER: JP 11-87192  
;; PRIOR FILING DATE: 1999-03-29  
;; NUMBER OF SEQ ID NOS: 36  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO: 14  
;; LENGTH: 116  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
US-09-966-880A-14

Query Match 13.1%; Score 78; DB 9; Length 116;  
Best Local Similarity 100.0%; Pred. No. 3.7e-32;  
Matches 78; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 466 GAAAGACTTTCAACCTCGGAGAGGCTGCATGAAAATTCAGTCTTCTCCAGACG 525  
DB 39 GAAAGACTTTCAACCTCGGAGAGGCTGCATGAAAATTCAGTCTTCTCCAGACG 98  
QY 526 CTTCGGCGCATCCTTTTG 543  
DB 99 CTTCGGCGCATCCTTTTG 116

RESULT 7  
US-09-966-880A-15

;; Sequence 15, Application US/09966880A  
;; Patent No. US20020164743A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Honjo, Tasuku  
;; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
;; FILE REFERENCE: 06501-088001  
;; CURRENT APPLICATION NUMBER: US/09/966, 880A  
;; PRIOR FILING DATE: 2001-09-28  
;; PRIOR APPLICATION NUMBER: PCT/JP00/01918  
;; PRIOR FILING DATE: 2000-03-28  
;; PRIOR APPLICATION NUMBER: JP 11-371382  
;; PRIOR FILING DATE: 1999-12-27  
;; PRIOR APPLICATION NUMBER: JP 11-178999  
;; PRIOR FILING DATE: 1999-06-24  
;; PRIOR APPLICATION NUMBER: JP 11-87192  
;; PRIOR FILING DATE: 1999-03-29  
;; NUMBER OF SEQ ID NOS: 36  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO: 15  
;; LENGTH: 2172  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
US-09-966-880A-15



Query Match 9.0%; Score 54; DB 9; Length 2172;  
Best Local Similarity 100.0%; Pred. No. 6.1e-19;  
Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 544 CCCCCTGATGAGCTGATGACTTACGAGACGCACTTCTGACTTGGGACTTGA 597  
Db 1 CCCCCTGATGAGCTGATGACTTACGAGACGCACTTCTGACTTGGGACTTGA 54

RESULT 8  
US-09-966-880A-1  
; Sequence 1, Application US/09966880A  
; Patent No. US20020164743A1

; GENERAL INFORMATION:  
; APPLICANT: Honjo, Tasuku  
; APPLICANT: Muramatsu, Masamichi  
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
; FILE REFERENCE: 06501-088001  
; CURRENT APPLICATION NUMBER: US/09/966,880A  
; PCT FILING DATE: 2001-09-28  
; PRIOR APPLICATION NUMBER: PCT/JP00/01918  
; PRIOR FILING DATE: 2000-03-28  
; PRIOR APPLICATION NUMBER: JP 11-371382  
; PRIOR FILING DATE: 1999-12-27  
; PRIOR APPLICATION NUMBER: JP 11-178999  
; PRIOR FILING DATE: 1999-06-24  
; PRIOR APPLICATION NUMBER: JP 11-87192  
; PRIOR FILING DATE: 1999-03-29  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 2440  
; TYPE: DNA  
; ORGANISM: Mus musculus  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (93)...(686)  
; NAME/KEY: 5' UTR  
; LOCATION: (1)...(92)  
; FEATURE:  
; NAME/KEY: 3' UTR  
; LOCATION: (690)...(2440)  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)...(2440)  
; OTHER INFORMATION: n = A, T, C or G  
US-09-966-880A-1

Query Match 5.4%; Score 32; DB 9; Length 2440;  
Best Local Similarity 100.0%; Pred. No. 7.5e-07;  
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 226 TACCGGTCACCTGCTTACCTCTGAGGCC 257  
Db 318 TACCGGTCACCTGCTTACCTCTGAGGCC 349

RESULT 9  
US-09-966-880A-18  
; Sequence 18, Application US/09966880A  
; Patent No. US20020164743A1  
; GENERAL INFORMATION:  
; APPLICANT: Honjo, Tasuku  
; APPLICANT: Muramatsu, Masamichi  
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
; FILE REFERENCE: 06501-088001  
; CURRENT APPLICATION NUMBER: US/09/966,880A  
; PCT FILING DATE: 2001-09-28  
; PRIOR APPLICATION NUMBER: PCT/JP00/01918  
; PRIOR FILING DATE: 2000-03-28  
; PRIOR APPLICATION NUMBER: JP 11-371382  
; PRIOR FILING DATE: 1999-12-27

; PRIOR APPLICATION NUMBER: JP 11-178999  
; PRIOR FILING DATE: 1999-06-24  
; PRIOR APPLICATION NUMBER: JP 11-87192  
; PRIOR FILING DATE: 1999-03-29  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 18  
; LENGTH: 30  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Artificially synthesized primer sequence, 22  
US-09-966-880A-18

Query Match 5.0%; Score 30; DB 9; Length 30;  
Best Local Similarity 100.0%; Pred. No. 8.5e-06;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 94 GTAGTGAAGAGCGCTGACAGTCTACATCC 123  
Db 1 GTAGTGAAGAGCGCTGACAGTCTACATCC 30

RESULT 10  
US-09-966-880A-24  
; Sequence 24, Application US/09966880A  
; Patent No. US20020164743A1  
; GENERAL INFORMATION:  
; APPLICANT: Honjo, Tasuku  
; APPLICANT: Muramatsu, Masamichi  
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
; FILE REFERENCE: 06501-088001  
; CURRENT APPLICATION NUMBER: US/09/966,880A  
; PCT FILING DATE: 2001-09-28  
; PRIOR APPLICATION NUMBER: PCT/JP00/01918  
; PRIOR FILING DATE: 2000-03-28  
; PRIOR APPLICATION NUMBER: JP 11-371382  
; PRIOR FILING DATE: 1999-12-27  
; PRIOR APPLICATION NUMBER: JP 11-178999  
; PRIOR FILING DATE: 1999-06-24  
; PRIOR APPLICATION NUMBER: JP 11-87192  
; PRIOR FILING DATE: 1999-03-29  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 24  
; LENGTH: 30  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Artificially synthesized primer sequence, p14  
US-09-966-880A-24

Query Match 5.0%; Score 30; DB 9; Length 30;  
Best Local Similarity 100.0%; Pred. No. 8.5e-06;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 AGCCTTGTGATGAACCGAGAGATTCTT 36  
Db 1 AGCCTTGTGATGAACCGAGAGATTCTT 30

RESULT 11  
US-09-966-880A-25/c  
; Sequence 25, Application US/09966880A  
; Patent No. US20020164743A1  
; GENERAL INFORMATION:  
; APPLICANT: Honjo, Tasuku  
; APPLICANT: Muramatsu, Masamichi  
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
; FILE REFERENCE: 06501-088001  
; CURRENT APPLICATION NUMBER: US/09/966,880A  
; PCT FILING DATE: 2001-09-28  
; PRIOR APPLICATION NUMBER: PCT/JP00/01918



;; PRIOR FILING DATE: 2000-03-28  
;; PRIOR APPLICATION NUMBER: JP 11-371382  
;; PRIOR FILING DATE: 1999-12-27  
;; PRIOR APPLICATION NUMBER: JP 11-178999  
;; PRIOR FILING DATE: 1999-06-24  
;; PRIOR APPLICATION NUMBER: JP 11-87192  
;; PRIOR FILING DATE: 1999-03-29  
;; NUMBER OF SEQ ID NOS: 36  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 25  
;; LENGTH: 28  
;; TYPE: DNA  
;; ORGANISM: Artificial Sequence  
;; FEATURE:  
;; OTHER INFORMATION: Artificially synthesized primer sequence, p16  
US-09-966-880A-25

Query Match 4.7%; Score 28; DB 9; Length 28;  
Best Local Similarity 100.0%; Pred. No. 0.00011;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 128 CACTGACCTTGGTATCTTCCCAATA 155  
DB 28 CACTGACCTTGGTATCTTCCCAATA 1

RESULT 12  
US-09-966-880A-28/c

;; Sequence 28, Application US/09966880A  
;; Patent No. US20020164743A1  
;; GENERAL INFORMATION:

;; APPLICANT: Muramatsu, Masamichi  
;; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
;; FILE REFERENCE: 06501-088001  
;; CURRENT APPLICATION NUMBER: US/09/966,880A  
;; CURRENT FILING DATE: 2001-09-28  
;; PRIOR APPLICATION NUMBER: PCT/JP00/01918  
;; PRIOR FILING DATE: 2000-03-28  
;; PRIOR APPLICATION NUMBER: JP 11-371382  
;; PRIOR FILING DATE: 1999-12-27  
;; PRIOR APPLICATION NUMBER: JP 11-178999  
;; PRIOR FILING DATE: 1999-06-24  
;; PRIOR APPLICATION NUMBER: JP 11-87192  
;; PRIOR FILING DATE: 1999-03-29  
;; NUMBER OF SEQ ID NOS: 36  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 28  
;; LENGTH: 28  
;; TYPE: DNA  
;; ORGANISM: Artificial Sequence  
;; FEATURE:  
;; OTHER INFORMATION: Artificially synthesized primer sequence, p26  
US-09-966-880A-28

Query Match 4.7%; Score 28; DB 9; Length 28;  
Best Local Similarity 100.0%; Pred. No. 0.00011;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 516 CTCGACAGCTTCGGCGATCCCTTTG 543  
DB 28 CTCGACAGCTTCGGCGATCCCTTTG 1

RESULT 13

US-09-966-880A-19/c  
;; Sequence 19, Application US/09966880A  
;; Patent No. US20020164743A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Honjo, Tasuku  
;; APPLICANT: Muramatsu, Masamichi  
;; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
;; FILE REFERENCE: 06501-088001

;; CURRENT APPLICATION NUMBER: US/09/966,880A  
;; CURRENT FILING DATE: 2001-09-28  
;; PRIOR APPLICATION NUMBER: PCT/JP00/01918  
;; PRIOR FILING DATE: 2000-03-28  
;; PRIOR APPLICATION NUMBER: JP 11-371382  
;; PRIOR FILING DATE: 1999-12-27  
;; PRIOR APPLICATION NUMBER: JP 11-178999  
;; PRIOR FILING DATE: 1999-06-24  
;; PRIOR APPLICATION NUMBER: JP 11-87192  
;; PRIOR FILING DATE: 1999-03-29  
;; NUMBER OF SEQ ID NOS: 36  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 19  
;; LENGTH: 27  
;; TYPE: DNA  
;; ORGANISM: Artificial Sequence  
;; FEATURE:  
;; OTHER INFORMATION: Artificially synthesized primer sequence, 25  
US-09-966-880A-19

Query Match 4.5%; Score 27; DB 9; Length 27;  
Best Local Similarity 100.0%; Pred. No. 0.00038;  
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 277 CATGTGCGCAGCTTCTGCGAGGAGAC 303  
DB 27 CATGTGCGCAGCTTCTGCGAGGAGAC 1

RESULT 14

US-09-918-995-30237  
;; Sequence 30237, Application US/09918995  
;; Publication No. US20030073623A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Hyseq, Inc.  
;; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
;; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES  
;; FILE REFERENCE: 20411-756  
;; CURRENT APPLICATION NUMBER: US/09/918,995  
;; CURRENT FILING DATE: 2001-07-30  
;; PRIOR APPLICATION NUMBER: US/09/235,076  
;; PRIOR FILING DATE: 1999-01-20  
;; NUMBER OF SEQ ID NOS: 38054  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 30237  
;; LENGTH: 371  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
US-09-918-995-30237

Query Match 3.7%; Score 22; DB 10; Length 371;  
Best Local Similarity 100.0%; Pred. No. 0.23;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 241 TTACCTCTGTGAGCCCTGCT 262  
DB 254 TTACCTCTGTGAGCCCTGCT 275

20NTS

RESULT 15

US-09-796-692-4504  
;; Sequence 4504, Application US/09796692  
;; Publication No. US20020198362A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Galger, Alexander  
;; APPLICANT: Maimon, Jane  
;; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERA  
;; TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES  
;; FILE REFERENCE: 2077.001200  
;; CURRENT APPLICATION NUMBER: US/09/796,692  
;; CURRENT FILING DATE: 2001-03-01  
;; PRIOR APPLICATION NUMBER: 60/186,126



```

; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO: 4504
; LENGTH: 553
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; LOCATION: (107)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (153)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (184)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (193)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (325)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (327)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (473)
; OTHER INFORMATION: n=A,T,C or G
; US-09-796-692-4504

Query Match      3.7%; Score 22; DB 9; Length 553;
Best Local Similarity 100.0%; Pred.No. 0.23;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY      241 TTCACTCTCGAGCCCTGCT 262
Db      531 TTCACTCTCGAGCCCTGCT 552
```

```

RESULT 16
US-10-040-862-4504
; Sequence 4504, Application US/10040862
; Publication No. US20030078396A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Aigate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Reiter, Marc
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
```

```

; TITLE OF INVENTION: Hematological Malignancies
; FILE REFERENCE: 014058-013520US
; CURRENT APPLICATION NUMBER: US/10/040,862
; CURRENT FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: US 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/223,378
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: US 09/796,692
; PRIOR FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 10467
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO: 4504
; LENGTH: 553
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; LOCATION: (107)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (153)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (184)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (193)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (325)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (327)
; OTHER INFORMATION: n=A,T,C or G
; NAME/KEY: unsure
; LOCATION: (473)
; OTHER INFORMATION: n=A,T,C or G
; US-10-040-862-4504

Query Match      3.7%; Score 22; DB 14; Length 553;
Best Local Similarity 100.0%; Pred.No. 0.23;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY      241 TTCACTCTCGAGCCCTGCT 262
Db      531 TTCACTCTCGAGCCCTGCT 552
```



RESULT 17  
US-10-057-475B-4504  
; Sequence 4504, Application US/10057475B  
; Publication No. US20040002068A1  
; GENERAL INFORMATION:  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Mannion, Jane  
; APPLICANT: Clapper, Jonathan David  
; APPLICANT: Wang, Aijun  
; APPLICANT: Ordonez, Nadia  
; APPLICANT: Carter, Lauren  
; APPLICANT: McNeill, Patricia Dianne  
; APPLICANT: Corixa Corporation  
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy  
; FILE REFERENCE: 014058-014402US  
; CURRENT APPLICATION NUMBER: US/10/057,475B  
; CURRENT FILING DATE: 2002-01-22  
; PRIOR APPLICATION NUMBER: US 60/186,126  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: US 60/190,479  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: US 60/200,545  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: US 60/200,303  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 60/200,779  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 60/200,999  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: US 60/202,084  
; PRIOR FILING DATE: 2000-05-04  
; PRIOR APPLICATION NUMBER: US 60/206,201  
; PRIOR FILING DATE: 2000-05-22  
; PRIOR APPLICATION NUMBER: US 60/218,950  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: US 60/222,903  
; PRIOR FILING DATE: 2000-08-03  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 10979  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 4504  
; LENGTH: 553  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(553)  
; OTHER INFORMATION: n = G, A, C or T  
US-10-057-475B-4504

Query Match 3.7%; Score 22; DB 15; Length 553;  
Best Local Similarity 100.0%; Pred. No. 0.23;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 241 TTACCTCCTGGAGCCCTGCT 262  
DB 531 TTACCTCCTGGAGCCCTGCT 552

RESULT 18  
US-10-154-884B-4504  
; Sequence 4504, Application US/10154884B  
; Publication No. US20040005561A1  
; GENERAL INFORMATION:  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Mannion, Jane  
; APPLICANT: Ketter, Marc W.  
; APPLICANT: Corixa Corporation

; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Thera  
; FILE REFERENCE: 014058-013521US  
; CURRENT APPLICATION NUMBER: US/10/154,884B  
; CURRENT FILING DATE: 2002-05-23  
; PRIOR APPLICATION NUMBER: US 60/186,126  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: US 60/190,479  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: US 60/200,545  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: US 60/200,303  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 60/200,779  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 60/200,999  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: US 60/202,084  
; PRIOR FILING DATE: 2000-05-04  
; PRIOR APPLICATION NUMBER: US 60/206,201  
; PRIOR FILING DATE: 2000-05-22  
; PRIOR APPLICATION NUMBER: US 60/218,950  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: US 60/222,903  
; PRIOR FILING DATE: 2000-08-03  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 11290  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 4504  
; LENGTH: 553  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(553)  
; OTHER INFORMATION: n = G, A, C or T  
US-10-154-884B-4504

Query Match 3.7%; Score 22; DB 15; Length 553;  
Best Local Similarity 100.0%; Pred. No. 0.23;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 241 TTACCTCCTGGAGCCCTGCT 262  
DB 531 TTACCTCCTGGAGCCCTGCT 552

RESULT 19  
US-09-796-692-5061/C  
; Sequence 5061, Application US/09796692  
; Publication No. US20020198362A1  
; GENERAL INFORMATION:  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Mannion, Jane  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERA  
; FILE REFERENCE: 2077-001200  
; CURRENT APPLICATION NUMBER: US/09/796,692  
; CURRENT FILING DATE: 2001-03-01  
; PRIOR APPLICATION NUMBER: 60/186,126  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: 60/190,479  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: 60/200,545  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: 60/200,303  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/200,779  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/200,999  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: 60/202,084



```

; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5061
; LENGTH: 556
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-796-692-5061

Query Match      3.7%  Score 22; DB 9; Length 556;
Best Local Similarity 100.0%; Pred. No. 0.23;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      241 TTACCTCCTGGAGCCCTGCT 262
DB      26  TTACCTCCTGGAGCCCTGCT 5

RESULT 20
US-10-040-862-5061/c
; Sequence 5061, Application US/10040862
; Publication No. US20030078396A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; FILE REFERENCE: 014058-013520US
; CURRENT APPLICATION NUMBER: US/10/040,862
; CURRENT FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: US 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/223,378
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: US 09/796,692
; PRIOR FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 10467
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5061
; LENGTH: 556
```

```

; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-040-862-5061

Query Match      3.7%  Score 22; DB 14; Length 556;
Best Local Similarity 100.0%; Pred. No. 0.23;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      241 TTACCTCCTGGAGCCCTGCT 262
DB      26  TTACCTCCTGGAGCCCTGCT 5

RESULT 21
US-10-057-475B-5061/c
; Sequence 5061, Application US/10057475B
; Publication No. US20040002068A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Clapper, Jonathan David
; APPLICANT: Wang, Aljun
; APPLICANT: Ordenez, Nadia
; APPLICANT: Carter, Lauren
; APPLICANT: McNeill, Patricia Dianne
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; FILE REFERENCE: 014058-014402US
; CURRENT APPLICATION NUMBER: US/10/057,475B
; CURRENT FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; Remaining Prior Application data removed - See file Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 10979
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5061
; LENGTH: 556
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-057-475B-5061

Query Match      3.7%  Score 22; DB 15; Length 556;
Best Local Similarity 100.0%; Pred. No. 0.23;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      241 TTACCTCCTGGAGCCCTGCT 262
DB      26  TTACCTCCTGGAGCCCTGCT 5

RESULT 22
US-10-154-884B-5061/c
; Sequence 5061, Application US/10154884B
```



Publication No. US20040005561A1  
GENERAL INFORMATION:  
APPLICANT: Gaiger, Alexander  
APPLICANT: Algate, Paul A.  
APPLICANT: Mannion, Jane  
APPLICANT: Retter, Marc W.  
TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy  
TITLE OF INVENTION: Hematological Malignancies  
FILE REFERENCE: 014058-013521US  
CURRENT APPLICATION NUMBER: US/10/154,884B  
CURRENT FILING DATE: 2002-05-23  
PRIOR APPLICATION NUMBER: US 60/186,126  
PRIOR FILING DATE: 2000-03-01  
PRIOR APPLICATION NUMBER: US 60/190,479  
PRIOR FILING DATE: 2000-03-17  
PRIOR APPLICATION NUMBER: US 60/200,545  
PRIOR FILING DATE: 2000-04-27  
PRIOR APPLICATION NUMBER: US 60/200,303  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: US 60/200,779  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: US 60/200,999  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: US 60/202,084  
PRIOR FILING DATE: 2000-05-04  
PRIOR APPLICATION NUMBER: US 60/206,201  
PRIOR FILING DATE: 2000-05-22  
PRIOR APPLICATION NUMBER: US 60/218,950  
PRIOR FILING DATE: 2000-07-14  
PRIOR APPLICATION NUMBER: US 60/222,903  
PRIOR FILING DATE: 2000-08-03  
Remaining prior application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 11290  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 5061  
LENGTH: 556  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-154-884B-5061

Query Match 3.7% Score 22; DB 15; Length 556;  
Best Local Similarity 100.0%; Pred. No. 0.23;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 241 TTCACCTCCTGGAGCCCTGCT 262  
Db 26 TTCACCTCCTGGAGCCCTGCT 5

RESULT 23  
US-10-460-923-1  
Sequence 1, Application US/10460923  
Publication No. US20040009951A1  
GENERAL INFORMATION:  
APPLICANT: MALIM, Michael H.  
APPLICANT: SHEEHY, Ann M.  
APPLICANT: HARRIS, Reuben S.  
APPLICANT: BISHOP, Kate N.  
APPLICANT: NEUBERGER, Michael S.  
APPLICANT: GADDIS, Nathan C.  
APPLICANT: SIMON, James H.M.  
TITLE OF INVENTION: DNA Deamination Mediates Innate Immunity to Retroviral Infection  
FILE REFERENCE: 22253-74380  
CURRENT APPLICATION NUMBER: US/10/460,923  
CURRENT FILING DATE: 2003-06-13  
PRIOR APPLICATION NUMBER: US 60/388,513  
PRIOR FILING DATE: 2002-06-13  
PRIOR APPLICATION NUMBER: US 60/472,952  
PRIOR FILING DATE: 2003-05-23  
NUMBER OF SEQ ID NOS: 12  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 1

LENGTH: 1155  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-460-923-1

Query Match 3.7% Score 22; DB 15; Length 1155;  
Best Local Similarity 100.0%; Pred. No. 0.23;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 241 TTCACCTCCTGGAGCCCTGCT 262  
Db 844 TTCACCTCCTGGAGCCCTGCT 865

RESULT 24  
US-09-729-674-173  
Sequence 173, Application US/09729674  
Patent No. US2001003935A1  
GENERAL INFORMATION:  
APPLICANT: Jacobs, Kenneth  
APPLICANT: McCoy, John M.  
APPLICANT: Lavalie, Edward R.  
APPLICANT: Collins-Racie, Lisa A.  
APPLICANT: Evans, Cheryl  
APPLICANT: Werberg, David  
APPLICANT: Treacy, Maurice  
APPLICANT: Agostino, Michael J.  
APPLICANT: Steinhilber II, Robert J.  
APPLICANT: Spaulding, Vikki  
APPLICANT: Wong, Gordon G.  
APPLICANT: Clark, Hilary  
APPLICANT: Rechtel, Kim  
APPLICANT: Genetics Institute, Inc.  
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM  
FILE REFERENCE: 6055-64X  
CURRENT APPLICATION NUMBER: US/09/729,674  
CURRENT FILING DATE: 2000-12-04  
PRIOR APPLICATION NUMBER: 09/539,330  
PRIOR FILING DATE: 2000-03-30  
NUMBER OF SEQ ID NOS: 283  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 173  
LENGTH: 1534  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-729-674-173

Query Match 3.7% Score 22; DB 9; Length 1534;  
Best Local Similarity 100.0%; Pred. No. 0.23;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 241 TTCACCTCCTGGAGCCCTGCT 262  
Db 894 TTCACCTCCTGGAGCCCTGCT 915

RESULT 25  
US-09-822-830A-359  
Sequence 359, Application US/09822830A  
Patent No. US20020142952A1  
GENERAL INFORMATION:  
APPLICANT: Genetics Institute, Inc.  
APPLICANT: Wong, Gordon G.  
APPLICANT: Clark, Hilary  
APPLICANT: Rechtel, Kim  
APPLICANT: Agostino, Michael J.  
APPLICANT: Howes, Steven H.  
APPLICANT: Resnick, Richard J.  
APPLICANT: Gulkota, Kamalakr  
APPLICANT: Graham, James R.  
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS  
FILE REFERENCE: GIN 6402  
CURRENT APPLICATION NUMBER: US/09/822,830A



Mon Mar 15 09:28:00 2004

us-09-966-880a-7\_copy\_80\_676.olg20.rmpb

Page 10

;/ CURRENT FILING DATE: 2001-03-29  
;/ PRIOR APPLICATION NUMBER: 60/195,604  
;/ PRIOR FILING DATE: 2000-04-06  
;/ NUMBER OF SEQ ID NOS: 631  
;/ SOFTWARE: PatentIn Ver. 2.0  
;/ SEQ ID NO 359  
;/ LENGTH: 2151  
;/ TYPE: DNA  
;/ ORGANISM: Homo sapiens  
US-09-822-830A-359

Query Match 3.7%; Score 22; DB 9; Length 2151;  
Best Local Similarity 100.0%; Pred. No. 0.24;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 241 TTCACCTCCTGGAGCCCTGCT 262  
DB 1533 TTCACCTCCTGGAGCCCTGCT 1554

RESULT 26  
US-09-966-880A-26  
;/ Sequence 26, Application US/09966880A  
;/ Patent No. US20020164743A1  
;/ GENERAL INFORMATION:  
;/ APPLICANT: Honjo, Tasuku  
;/ APPLICANT: Muramatsu, Masamichi  
;/ TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
;/ FILE REFERENCE: 06501-088001  
;/ CURRENT APPLICATION NUMBER: US/09/966,880A  
;/ CURRENT FILING DATE: 2001-09-28  
;/ PRIOR APPLICATION NUMBER: PCT/JP00/01918  
;/ PRIOR FILING DATE: 2000-03-28  
;/ PRIOR APPLICATION NUMBER: JP 11-371382  
;/ PRIOR FILING DATE: 1999-12-27  
;/ PRIOR APPLICATION NUMBER: JP 11-178999  
;/ PRIOR FILING DATE: 1999-06-24  
;/ PRIOR APPLICATION NUMBER: JP 11-87192  
;/ PRIOR FILING DATE: 1999-03-29  
;/ NUMBER OF SEQ ID NOS: 36  
;/ SOFTWARE: FastSeq for Windows Version 4.0  
;/ SEQ ID NO 26  
;/ LENGTH: 21  
;/ TYPE: DNA  
;/ ORGANISM: Artificial Sequence  
;/ FEATURE:  
;/ OTHER INFORMATION: Artificially synthesized primer sequence, p17  
US-09-966-880A-26

Query Match 3.5%; Score 21; DB 9; Length 21;  
Best Local Similarity 100.0%; Pred. No. 0.75;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 209 TAGACCTGCGCGCTGCTACC 229  
DB 1 TAGACCTGCGCGCTGCTACC 21

RESULT 27  
US-10-292-198-1  
;/ Sequence 1, Application US/10292198  
;/ Publication No. US20030157654A1  
;/ GENERAL INFORMATION:  
;/ APPLICANT: SHEN, Ben  
;/ APPLICANT: LIU, Wen  
;/ TITLE OF INVENTION: BIOSYNTHEIS OF ENEDIYNE COMPOUNDS BY MANIPULATION OF C-1027 GENE  
;/ FILE REFERENCE: 054030-0007  
;/ CURRENT APPLICATION NUMBER: US/10/292,198  
;/ CURRENT FILING DATE: 2003-03-14  
;/ PRIOR APPLICATION NUMBER: US 10/159,257  
;/ PRIOR FILING DATE: 2002-05-31  
;/ PRIOR APPLICATION NUMBER: US 09/478,188

;/ PRIOR FILING DATE: 2000-01-05  
;/ PRIOR APPLICATION NUMBER: US 60/115,434  
;/ PRIOR FILING DATE: 1999-01-06  
;/ NUMBER OF SEQ ID NOS: 146  
;/ SOFTWARE: PatentIn version 3.2  
;/ SEQ ID NO 1  
;/ LENGTH: 63158  
;/ TYPE: DNA  
;/ ORGANISM: Streptomyces globisporus  
US-10-292-198-1

Query Match 3.4%; Score 20; DB 14; Length 63158;  
Best Local Similarity 100.0%; Pred. No. 3.2;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 368 CCGAGGGGCTGCGCGGCTG 387  
DB 55944 CCGAGGGGCTGCGCGGCTG 55963

Search completed: March 12, 2004, 17:42:55  
Job time : 195 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 12, 2004, 09:38:52 ; Search time 65 Seconds

(without alignments)  
5097.011 Million cell updates/sec

Title: US-09-966-880A-7\_COPY\_80\_676

Perfect score: 597  
Sequence: 1 atgacagacccttcgatgaa.....ctcgacttcggacttga 597

Scoring table: OLIGO.MVC  
Gapop 60.0 , Gapext 60.0

Searched: 682709 seqs, 27747546 residues

Word size: 20

Total number of hits satisfying chosen parameters: 2

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database: Issued Patents NA:

1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	20	3.4	1740	2 US-08-796-488-1	Sequence 1, Appli
C 2	20	3.4	1740	3 US-09-243-934-1	Sequence 1, Appli

## ALIGNMENTS

RESULT 1  
US-08-796-488-1/c  
Sequence 1, Application US/08796488  
Patent No. 5928927  
GENERAL INFORMATION:  
APPLICANT: CHENG, TU-CHEN  
APPLICANT: DEFRANK, JOSEPH J.  
TITLE OF INVENTION: ENZYMATIC DETOXIFICATION OF  
TITLE OF INVENTION: ORGANOPOPHORUS COMPOUNDS  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: U.S. Army Chemical and Biological Defense  
ADDRESSER: Command  
STREET: Office of the Chief Counsel, Bldg E4435  
CITY: Aberdeen Proving Ground  
STATE: MD  
COUNTRY: U.S.  
ZIP: 21010-5423  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

102e  
DIV of  
015,966  
116/96  
Priority

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/796,488  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Bilton, U. J.  
REGISTRATION NUMBER: 39,908  
REFERENCE/DOCKET NUMBER: DAM 443-96  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 410-671-1158  
TELEFAX: 410-671-2534  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1740 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Alteromonas sp.  
STRAIN: JD6.5  
IMMEDIATE SOURCE:  
CLONE: pTC6513  
FEATURE:  
NAME/KEY: -10 signal  
LOCATION: 14..18  
FEATURE:  
NAME/KEY: RBS  
LOCATION: 24..30  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 36..1586  
US-08-796-488-1

Query Match 3.4%; Score 20; DB 2; Length 1740;  
Best Local Similarity 100.0%; Pred. No. 0.84;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 245 CCTCTGAGACCCCTGCTAC 264 20NT  
Db 1605 CCTCTGAGACCCCTGCTAC 1586

RESULT 2  
US-09-243-934-1/c  
Sequence 1, Application US/09243934  
Patent No. 6080566  
GENERAL INFORMATION:  
APPLICANT: CHENG, TU-CHEN  
APPLICANT: DEFRANK, JOSEPH J.  
TITLE OF INVENTION: ENZYMATIC DETOXIFICATION OF  
TITLE OF INVENTION: ORGANOPOPHORUS COMPOUNDS  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: U.S. Army Chemical and Biological Defense  
ADDRESSER: Command  
STREET: Office of the Chief Counsel, Bldg E4435  
CITY: Aberdeen Proving Ground  
STATE: MD  
COUNTRY: U.S.  
ZIP: 21010-5423  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/243,934



```

/ FILING DATE:
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/796,488
/ FILING DATE:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Biffoni, U. J.
/ REGISTRATION NUMBER: 39,908
/ REFERENCE/DOCKET NUMBER: DAM 443-96
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 410-671-1158
/ TELEFAX: 410-671-2534
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1740 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ HYPOTHETICAL: NO
/ ANTI-SENSE: NO
/ ORIGINAL SOURCE:
/ ORGANISM: Alteromonas sp.
/ STRAIN: JDE.5
/ IMMEDIATE SOURCE:
/ CLONE: pTC6513
/ FEATURE:
/ NAME/KEY: -10_signal
/ LOCATION: 14..18
/ FEATURE:
/ NAME/KEY: RBS
/ LOCATION: 24..30
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 36..1586
/ US-09-243-934-1

```

```

Query Match      3.4%; Score 20; DB 3; Length 1740;
Best Local Similarity 100.0%; Pred. No. 0.84;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      245 CCTCTGAGAGCCCTGCTAC 264
Db      1605 CCTCTGAGAGCCCTGCTAC 1586

```

Search completed: March 12, 2004, 12:27:53  
Job time : 65 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 13, 2004, 05:04:59 ; Search time 2369 Seconds

(without alignments)  
1096.669 Million cell updates/sec

Title: US-09-966-880A-11

Sequence: 1 agagagaccatcatatcga.....ctggagaccactatgacag 87

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 20

Total number of hits satisfying chosen parameters: 17

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 100 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hcc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hcc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pin:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fam:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	87	100.0	535	14	CD707143
2	87	100.0	693	12	EG757392
3	87	100.0	1201	13	BX402063
4	84	96.6	541	10	BF238155

5	84	96.6	743	12	BG686133	602638412
6	84	96.6	942	10	BF975166	602244657
7	71	81.6	953	13	B0065440	AGENCOURT
8	71	81.6	1052	13	B0055935	AGENCOURT
9	51	58.6	820	12	BG757089	602715124
10	37	42.5	889	12	BG686876	602650861
11	36	41.4	872	12	BG758510	602712721
12	34	39.1	1201	9	AL559877	AL559877
13	26	29.9	853	13	BX464579	BX464579
14	21	24.1	730	28	AQ781292	HS_3104_B
15	20	23.0	520	9	A1936315	w064909.X
16	20	23.0	617	13	BU200032	604152068
17	20	23.0	701	13	BU351369	6035829226

## ALIGNMENTS

RESULT 1  
LOCUS CD707143 535 bp mRNA linear EST 25-JUN-2003  
DEFINITION EST23670 human nasopharynx Homo sapiens cDNA, mRNA sequence.  
ACCESSION CD707143  
VERSION CD707143.1 GI:32237773  
KEYWORDS EST.

SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE Liu,X.-Q., Zhou,Y., Zhang,L.-J., Xu,H., Chen,H.-K., Pan,Z.-G. and Zeng,Y.-X.  
Transcriptional Gene Expression Profile of Human Nasopharynx  
JOURNAL Unpublished (2003)  
COMMENT Contact: Yixun Zeng  
Cancer Center

AUTHORS Sun Yat-sen University  
651 Dongfeng Road East, Guangzhou 510060, China  
Tel: 86-1380-9770-743  
Fax: 86-20-8775-4506  
Email: yxzeng@qzsun.edu.cn.

FEATURES  
Source  
1..535  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/issue\_type="normal nasopharynx"  
/clone\_lib="human nasopharynx"  
/note="ESTs generated from a normal nasopharynx cDNA library from southern Chinese"

## ORIGIN

Query Match 100.0%; Score 87; DB 14; Length 535;  
Best Local Similarity 100.0%; Pred. No. 4, 5e-37;

Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	AGAGAACCATCATTAATTAAGTGAATTTTCTGACCTGAGACTTGACGAGGAGCAGAGA	60
DB <td>42</td> <td>AGAGAACCATCATTAATTAAGTGAATTTTCTGACCTGAGACTTGACGAGGAGCAGAGA</td> <td>101</td>	42	AGAGAACCATCATTAATTAAGTGAATTTTCTGACCTGAGACTTGACGAGGAGCAGAGA	101
QY <td>61</td> <td>AGACACTCTGGACACCACTATGACAG<td>87</td></td>	61	AGACACTCTGGACACCACTATGACAG <td>87</td>	87
DB <td>102</td> <td>AGACACTCTGGACACCACTATGACAG</td> <td>128</td>	102	AGACACTCTGGACACCACTATGACAG	128

RESULT 2  
LOCUS EG757392 693 bp mRNA linear EST 15-MAY-2001  
DEFINITION 602711022F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:485180 5',  
mRNA sequence.  
ACCESSION EG757392  
VERSION EG757392.1 GI:14068045  
KEYWORDS EST.



**SOURCE** Homo sapiens (human)  
**ORGANISM** Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

**REFERENCE** 1 (bases 1 to 693)  
 NIH-MGC <http://mgi.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgaabp-remail.nih.gov  
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>  
 Plate: LNCM1694 row: K column: 05  
 High quality sequence stop: 693.  
 Location/Qualifiers  
 1..693  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4851580"  
 /tissue\_type="primary B-cells from tonsils (cell line)"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_1lb="NIH\_MGC\_48"  
 /note="Organ: B-cells; Vector: pOTB7; Site: 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC library."

**ORIGIN**  
 Query Match 100.0%; Score 87; DB 12; Length 693;  
 Best Local Similarity 100.0%; Pred. No. 4.7e-37;  
 Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

**QY** 1 AGAGAACCATCATTAATTGAGTGAATTTCTGCGCTGAGACTTGACGGAGGCAAGA 60  
 |||||||  
**DB** 2 AGAGAACCATCATTAATTGAGTGAATTTCTGCGCTGAGACTTGACGGAGGCAAGA 61  
 |||||||  
**QY** 61 AGAGACTCTGGACACCATATGACAG 87  
 |||||||  
**DB** 62 AGAGACTCTGGACACCATATGACAG 88  
 |||||||

**RESULT 3**  
**LOCUS** BX402063 1201 bp mRNA linear EST 13-MAY-2003  
**DEFINITION** BX402063 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED  
 Homo sapiens cDNA clone CSODL012YD18 5-PRIME, mRNA sequence.  
**ACCESSION** BX402063  
**VERSION** BX402063.1 GI:30626645  
**KEYWORDS** EST.  
**SOURCE** Homo sapiens (human)  
**ORGANISM** Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

**REFERENCE** 1 (bases 1 to 1201)  
 Li, W.B., Gruber, C., Tessee, J. and Polayes, D.  
 Full-length cDNA libraries and normalization  
 Unpublished (2001)  
 Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 Evry cedex - France  
 Email: seqref@genoscope.cns.fr; Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
 Library was constructed by Life Technologies, a division of

**FEATURES**  
 source  
 1..1201  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CSODL012YD18"  
 /tissue\_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"  
 /cell\_line="RAMOS CELL LINE"  
 /clone\_1lb="Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

**ORIGIN**  
 Query Match 100.0%; Score 87; DB 13; Length 1201;  
 Best Local Similarity 100.0%; Pred. No. 5e-37;  
 Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

**QY** 1 AGAGAACCATCATTAATTGAGTGAATTTCTGCGCTGAGACTTGACGGAGGCAAGA 60  
 |||||||  
**DB** 73 AGAGAACCATCATTAATTGAGTGAATTTCTGCGCTGAGACTTGACGGAGGCAAGA 132  
 |||||||  
**QY** 61 AGAGACTCTGGACACCATATGACAG 87  
 |||||||  
**DB** 133 AGAGACTCTGGACACCATATGACAG 159  
 |||||||

**RESULT 4**  
**LOCUS** BF238155 541 bp mRNA linear EST 14-NOV-2000  
**DEFINITION** 601811880F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4054915 5', mRNA sequence.  
**ACCESSION** BF238155  
**VERSION** BF238155.1 GI:1152074  
**KEYWORDS** EST.  
**SOURCE** Homo sapiens (human)  
**ORGANISM** Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

**REFERENCE** 1 (bases 1 to 541)  
 NIH-MGC <http://mgi.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgaabp-remail.nih.gov  
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>  
 Plate: LNCM895 row: P column: 20  
 High quality sequence stop: 541.  
 Location/Qualifiers  
 1..541  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4054915"  
 /tissue\_type="primary B-cells from tonsils (cell line)"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_1lb="NIH\_MGC\_48"  
 /note="Organ: B-cells; Vector: pOTB7; Site: 1: XhoI;



site 2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCAACGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by Ling  
Hong in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 96.6%; Score 84; DB 10; Length 541;  
Best Local Similarity 100.0%; Pred. No. 2e-35;  
Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GAACCATCATTAAATGAAGTGAATTTCTGCGCTGAGACTTGCAGGGAGGCAAGAGA 63  
DB 2 GAACCATCATTAAATGAAGTGAATTTCTGCGCTGAGACTTGCAGGGAGGCAAGAGA 61

QY 64 CACTCTGAGACCACTATGACAG 87  
DB 62 CACTCTGAGACCACTATGACAG 85

RESULT 5 743 bp mRNA linear EST 01-MAY-2001  
BG686133 602638412F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4766234 5',  
LOCUS mRNA sequence.

ACCESSION BG686133  
VERSION BG686133.1 GI:13917530  
KEYWORDS EST.

SOURCE Homo sapiens (human)  
ORGANISM

REFERENCE 1 (bases 1 to 743)  
AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.

cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
Plate: LNCM1626 row: 5 column: 03  
High quality sequence stop: 740.  
Location/Qualifiers

FEATURES  
source

1..743  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4766234"  
/issue\_type="primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_48"  
/note="Organ: B-cells; Vector: pOT7; Site\_1: XhoI;  
Site\_2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCAACGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by Ling  
Hong in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 96.6%; Score 84; DB 12; Length 743;  
Best Local Similarity 100.0%; Pred. No. 2.1e-35;  
Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GAACCATCATTAAATGAAGTGAATTTCTGCGCTGAGACTTGCAGGGAGGCAAGAGA 63  
DB 2 GAACCATCATTAAATGAAGTGAATTTCTGCGCTGAGACTTGCAGGGAGGCAAGAGA 61

QY 64 CACTCTGAGACCACTATGACAG 87  
DB 62 CACTCTGAGACCACTATGACAG 85

RESULT 6 942 bp mRNA linear EST 22-JAN-2001  
BP975166 602244657F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:435639 5',  
LOCUS mRNA sequence.

ACCESSION BP975166  
VERSION BP975166.1 GI:12342381  
KEYWORDS EST.

SOURCE Homo sapiens (human)  
ORGANISM

REFERENCE 1 (bases 1 to 942)  
AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.

cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
Plate: LNCM1207 row: a column: 16  
High quality sequence stop: 707.  
Location/Qualifiers

FEATURES  
source

1..942  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:435639"  
/issue\_type="primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_48"  
/note="Organ: B-cells; Vector: pOT7; Site\_1: XhoI;  
Site\_2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCAACGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by Ling  
Hong in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 96.6%; Score 84; DB 10; Length 942;  
Best Local Similarity 100.0%; Pred. No. 2.2e-35;  
Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GAACCATCATTAAATGAAGTGAATTTCTGCGCTGAGACTTGCAGGGAGGCAAGAGA 63  
DB 2 GAACCATCATTAAATGAAGTGAATTTCTGCGCTGAGACTTGCAGGGAGGCAAGAGA 61

QY 64 CACTCTGAGACCACTATGACAG 87  
DB 62 CACTCTGAGACCACTATGACAG 85

RESULT 7 953 bp mRNA linear EST 02-APR-2002  
BQ065440 BQ065440



DEFINITION AGENCOURT 6855061 NIH\_MGC\_99 Homo sapiens cDNA clone IMAGE:5929977  
5', mRNA sequence.  
ACCESSION B065440  
VERSION B065440.1 GI:19894486  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 953)  
NIH-MGC <http://mgs.nci.nih.gov/>.  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE Unpublished (1999)  
JOURNAL  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Lou Staudt  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>  
Plate: LNCM2108 row: P column: 10  
High quality sequence stop: 634.  
Location/Qualifiers

FEATURES  
source  
1..953  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5929977"  
/tissue\_type="lymphoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC\_99"  
/note="Organ: lymph; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC library."

ORIGIN  
Query Match 81.6%; Score 71; DB 13; Length 953;  
Best Local Similarity 100.0%; Pred. No. 3.3e-28;  
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 TTGAAGTGAATTTTCTGGCTGAGACTTGCAAGGAGGCAAGAGACACTCTGACACC 76  
1 TTGAAGTGAATTTTCTGGCTGAGACTTGCAAGGAGGCAAGAGACACTCTGACACC 60

DB 1 TTGAAGTGAATTTTCTGGCTGAGACTTGCAAGGAGGCAAGAGACACTCTGACACC 60

QY 77 ACTATGGACAG 87  
1 ACTATGGACAG 71

DB 61 ACTATGGACAG 71

RESULT 8  
LOCUS B0655935 1052 bp mRNA linear EST 29-MAR-2002  
DEFINITION AGENCOURT 6796291 NIH\_MGC\_99 Homo sapiens cDNA clone IMAGE:5808181  
5', mRNA sequence.  
ACCESSION B0655935  
VERSION B0655935.1 GI:19815262  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1052)  
NIH-MGC <http://mgs.nci.nih.gov/>.  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE Unpublished (1999)  
JOURNAL

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Lou Staudt  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>  
Plate: LNCM2051 row: W column: 14  
High quality sequence stop: 665.  
Location/Qualifiers

FEATURES  
source  
1..1052  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5808181"  
/tissue\_type="lymphoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC\_99"  
/note="Organ: lymph; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC library."

ORIGIN  
Query Match 81.6%; Score 71; DB 13; Length 1052;  
Best Local Similarity 100.0%; Pred. No. 3.3e-28;  
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 TTGAAGTGAATTTTCTGGCTGAGACTTGCAAGGAGGCAAGAGACACTCTGACACC 76  
1 TTGAAGTGAATTTTCTGGCTGAGACTTGCAAGGAGGCAAGAGACACTCTGACACC 60

DB 1 TTGAAGTGAATTTTCTGGCTGAGACTTGCAAGGAGGCAAGAGACACTCTGACACC 60

QY 77 ACTATGGACAG 87  
1 ACTATGGACAG 71

DB 61 ACTATGGACAG 71

RESULT 9  
LOCUS BG757089 820 bp mRNA linear EST 15-MAY-2001  
DEFINITION 602715124F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:485517 5', mRNA sequence.  
ACCESSION BG757089  
VERSION BG757089.1 GI:14067742  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 820)  
NIH-MGC <http://mgs.nci.nih.gov/>.  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE Unpublished (1999)  
JOURNAL  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>  
Plate: LNCM1704 row: O column: 06  
High quality sequence stop: 675.  
Location/Qualifiers

FEATURES  
source  
1..820  
Location/Qualifiers



/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:485517"  
/tissue\_type="Primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_id="NIH\_MGC\_48"  
/note="Organ: B-cells; Vector: pOTB7; Site\_1: XhoI;  
Site\_2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by Ling  
Hong in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 58.6%; Score 51; DB 12; Length 820;  
Best Local Similarity 100.0%; Pred. No. 3.5e-17;  
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGAGACCATCATTAATGAGTGAATTTCTGCGCTGAGACTGCAGG 51  
|||||  
6 AGAGACCATCATTAATGAGTGAATTTCTGCGCTGAGACTGCAGG 56

Db

RESULT 10 889 bp mRNA linear EST 01-MAY-2001  
BG686876 602650861F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4763247 5',  
LOCUS mRNA sequence.  
BG686876  
ACCESSION BG686876.1 GI:13918273  
VERSION  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE NIH-MGC http://mgc.nci.nih.gov/  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE Unpublished (1999)  
JOURNAL  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabs-remail.nih.gov  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: LNCM1618 row: j column: 16  
High quality sequence start: 6  
High quality sequence stop: 727.  
Location/Qualifiers

1.889  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4763247"  
/tissue\_type="Primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_id="NIH\_MGC\_48"  
/note="Organ: B-cells; Vector: pOTB7; Site\_1: XhoI;  
Site\_2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by Ling  
Hong in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."

## FEATURES

source

1.889  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4763247"  
/tissue\_type="Primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_id="NIH\_MGC\_48"  
/note="Organ: B-cells; Vector: pOTB7; Site\_1: XhoI;  
Site\_2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by Ling  
Hong in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 42.5%; Score 37; DB 12; Length 889;  
Best Local Similarity 100.0%; Pred. No. 1.9e-09;  
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 51 GGAGGCAAGAGACACTGTGACACACTATGACAG 87  
|||||  
Db 9 GGAGGCAAGAGACACTGTGACACACTATGACAG 45

RESULT 11 872 bp mRNA linear EST 15-MAY-2001  
BG758510 602712721F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4853069 5',  
LOCUS mRNA sequence.  
BG758510  
ACCESSION BG758510.1 GI:14069163  
VERSION  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE NIH-MGC http://mgc.nci.nih.gov/  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE Unpublished (1999)  
JOURNAL  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabs-remail.nih.gov  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: LNCM1698 row: i column: 06  
High quality sequence stop: 836.  
Location/Qualifiers

## FEATURES

source

1.872  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4853069"  
/tissue\_type="Primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_id="NIH\_MGC\_48"  
/note="Organ: B-cells; Vector: pOTB7; Site\_1: XhoI;  
Site\_2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by Ling  
Hong in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 41.4%; Score 36; DB 12; Length 872;  
Best Local Similarity 100.0%; Pred. No. 6.8e-09;  
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 16 ATTGAAGTGAATTTCTGCGCTGAGACTTGACAG 51  
|||||  
Db 2 ATTGAAGTGAATTTCTGCGCTGAGACTTGACAG 37

RESULT 12 1201 bp mRNA linear EST 31-MAY-2003  
AL559877 AL559877 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA  
LOCUS clone CSDDG003YB14 5-PRIME, mRNA sequence.  
DEFINITION  
ACCESSION AL559877



VERSION AL559877.2 GI:31284008  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 1201)  
 AUTHORS Li W.B., Gruber C., Jesse J., and Polayes D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT On Feb 15, 2001 this sequence version replaced gi:12905793.  
 Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 Evry cedex - France  
 Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 6672.r For more information about this cluster, see  
 http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0D003DA070P1&cluster=6672.r. Contact : Feng Liang Email : fliang@life-tech.com URL : http://fulllength.invitrogen.com/Invitrogen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0D003DA070P1.

FEATURES  
 source  
 1..1201  
 /organism="Homo sapiens"  
 /mol\_type="RNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0D003YB14"  
 /tissue\_type="B CELLS (RAMOS CELL LINE)"  
 /cell\_line="RAMOS CELL LINE"  
 /note="Homo sapiens B CELLS (RAMOS CELL LINE)"  
 /clone\_lib="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

ORIGIN  
 Query Match 39.1%; Score 34; DB 9; Length 1201;  
 Best Local Similarity 100.0%; Pred. No. 8.9e-08;  
 Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 54 GGCAGAGAGAGACTGTGACACCACTATGACAG 87  
 66 GGCAGAGAGAGACTGTGACACCACTATGACAG 99

RESULT 13  
 LOCUS BX464579 853 bp mRNA linear EST 22-MAY-2003  
 DEFINITION BX464579 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA  
 clone CS0D003YB14 5-PRIME, mRNA sequence.  
 ACCESSION BX464579  
 VERSION BX464579.1 GI:31031641  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 853)  
 AUTHORS Li W.B., Gruber C., Jesse J., and Polayes D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 Evry cedex - France  
 Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 6672.r For more information about this cluster, see  
 http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS1D001ZF10Q1&cluster=6672.r. Contact :

Feng Liang Email : fliang@life-tech.com URL : http://fulllength.invitrogen.com/Invitrogen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS1D001ZF10Q1.

FEATURES  
 source  
 1..853  
 /organism="Homo sapiens"  
 /mol\_type="RNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0D003YB14"  
 /tissue\_type="B CELLS (RAMOS CELL LINE)"  
 /cell\_line="RAMOS CELL LINE"  
 /note="Homo sapiens B CELLS (RAMOS CELL LINE)"  
 /clone\_lib="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

ORIGIN  
 Query Match 29.9%; Score 26; DB 13; Length 853;  
 Best Local Similarity 100.0%; Pred. No. 0.0022;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 62 GACACTGTGACACCACTATGACAG 87  
 69 GACACTGTGACACCACTATGACAG 94

RESULT 14  
 LOCUS AQ781292 730 bp DNA linear GSS 02-AUG-1999  
 DEFINITION HS 3104\_B1\_B04\_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3104 Col=7 Row=D, genomic survey sequence.  
 ACCESSION AQ781292  
 VERSION AQ781292.1 GI:5684252  
 KEYWORDS GSS.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 730)  
 AUTHORS Mahairas G.G., Wallace J.C., Smith K., Swartzell S., Holzman T., Keller A., Shaker R., Furlong J., Young J., Zhao S., Adams M.D. and Hood L.  
 TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
 MEDLINE 99380589  
 PUBMED 10449764  
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
 High Throughput Sequencing Center  
 University of Washington  
 401 Queen Anne Avenue North, Seattle, WA 98109, USA  
 Tel: (206) 616-3618  
 Fax: (206) 616-3887  
 Email: jwallace@u.washington.edu  
 Clones may be purchased from Research Genetics (info@resgen.com).  
 BAC and Web Server: http://www.htsc.washington.edu  
 Plate: 3104 row: D column: 7  
 Seq primer: M13 Reverse  
 Class: BAC ends  
 High quality sequence stop: 730.  
 Location/Qualifiers  
 1..730  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /clone="Plate=3104 Col=7 Row=D"  
 /sex="male"  
 /note="Organ: sperm; Vector: pBeloBAC11; BAC clones in E-Coli DH10B"



## ORIGIN

Query Match 24.1%; Score 21; DB 28; Length 730;  
 Best Local Similarity 100.0%; Pred. No. 1.3;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 51 GGAGGCAAGAGACACTCTCG 71  
 DB 68 GGAGGCAAGAGACACTCTCG 88

21 X

RESULT 15  
 AI936315 520 bp mRNA linear EST 08-MAR-2000  
 LOCUS 6064909.x1 NCI CGAP Pr23 Homo sapiens cDNA clone IMAGE:2460160 3'  
 DEFINITION similar to gb:U63488 REPLICATION PROTEIN A 70 KD DNA-BINDING  
 SUBUNIT (HUMAN); mRNA sequence.

ACCESSION AI936315 GI:5675185  
 VERSION  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 520)  
 NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 CONTACT: Robert Strausberg, Ph.D.  
 Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)  
 Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.  
 DNA Sequencing by: Greg Lennon, Ph.D.  
 Clone distribution: NCI-CCGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LINAC:  
[www-bio.livl.gov/bbtp/image/image.html](http://www-bio.livl.gov/bbtp/image/image.html)  
 Insert Length: 634 Std Error: 0.00  
 Seq primer: -40up from Gibco  
 High quality sequence stop: 444.

FEATURES  
 Source  
 1..520  
 location/Qualifiers

/organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2460160"  
 /sex="male"  
 /tissue\_type="normal prostate"  
 /lab\_host="DH10B"  
 /clone\_lib="NCI CGAP Pr22"  
 /note="Organ: prostate; Vector: pRT3D-Pac (Pharmacia)  
 with a modified polylinker; 1st strand cDNA was prepared  
 from normal prostate bulk tissue, and was then primed with  
 a Not I - oligo(dT) primer. Double-stranded cDNA was  
 ligated to Eco RI adaptors (Pharmacia), digested with Not  
 I and cloned into the Not I and Eco RI sites of the  
 modified pRT3 vector. Library is normalized, and was  
 constructed by Bento Soares and M. Fatima Bonaldo."

## ORIGIN

Query Match 23.0%; Score 20; DB 9; Length 520;  
 Best Local Similarity 100.0%; Pred. No. 4.3;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 CATCATTAATTGAGTGAGA 27  
 DB 159 CATCATTAATTGAGTGAGA 178

RESULT 16  
 BU200032/c

LOCUS BU200032 617 bp mRNA linear EST 25-NOV-2002  
 DEFINITION 604152068F1 CSECHN03 Gallus gallus cDNA clone CHEST994518 5', mRNA  
 sequence.

ACCESSION BU200032  
 VERSION BU200032.1 GI:25363376  
 KEYWORDS EST.  
 SOURCE Gallus gallus (chicken)  
 ORGANISM Gallus gallus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 Phasianinae; Gallus.

REFERENCE 1 (bases 1 to 617)  
 Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,  
 Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.  
 A Comprehensive Collection of Chicken cDNAs  
 Curr. Biol. 12 (22), 1965-1969 (2002)  
 JOURNAL MEDLINE  
 PUBMED 22335534  
 CONTACT: Simon Hubbard  
 Department of Biomolecular Sciences  
 University of Manchester Institute of Science and Technology  
 (UMIST)  
 PO Box 88, Manchester, M60 1QD, UK  
 Tel: 01612008930  
 Fax: 01612360409  
 Email: [Simon.Hubbard@umist.ac.uk](mailto:Simon.Hubbard@umist.ac.uk).

FEATURES  
 Source  
 1..617  
 location/Qualifiers

/organism="Gallus gallus"  
 /mol\_type="mRNA"  
 /strain="White Leghorn, Hisex"  
 /db\_xref="taxon:9031"  
 /clone="CHEST994918"  
 /tissue\_type="whole embryo"  
 /dev\_stage="20-21"  
 /lab\_host="DH10B"  
 /clone\_lib="CSECHN03"  
 /note="Organ: whole embryo; Vector: pluescript II KS(+);  
 site\_1: EcoRI; site\_2: NotI; This normalized library was  
 constructed from 1 million independent clones. cDNA  
 synthesis was initiated using an oligo(dT) primer, using  
 methylated C in the first strand synthesis reaction.  
 Following this first strand reaction, double-stranded cDNA  
 was bluntend, ligated to NotI adaptors, digested with  
 EcoRI, size-selected, and cloned into the NotI and EcoRI  
 compatible sites of a custom modified MCS of the  
 pluescript (KS+) vector. The library was normalized in 2  
 rounds using conditions adapted from Soares et al., PNAS  
 (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6  
 (1996): 791, except that a significantly longer  
 reannealing hybridization was used."

## ORIGIN

Query Match 23.0%; Score 20; DB 13; Length 617;  
 Best Local Similarity 100.0%; Pred. No. 4.4;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 TTATATGAGTGAGATTTT 32  
 DB 337 TTATATGAGTGAGATTTT 318

RESULT 17  
 BU351369/c 701 bp mRNA linear EST 28-NOV-2002  
 LOCUS 603529226F1 CSECHN69 Gallus gallus cDNA clone CHEST46067 5', mRNA  
 DEFINITION sequence.

ACCESSION BU351369  
 VERSION BU351369.1 GI:25859370  
 KEYWORDS EST.  
 SOURCE Gallus gallus (chicken)  
 ORGANISM Gallus gallus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;



Archosauria; Aves; Neognathae; Galliformes; Phasianidae;

Phasianinae; Gallus.

1 (bases 1 to 701)

Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,

Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.

A Comprehensive Collection of Chicken cDNAs

Curr. Biol. 12 (22), 1965-1969 (2002)

22335534

JOURNAL

MEDLINE

PUBMED

12445392

CONTACT: Simon Hubbard

Department of Biomolecular Sciences

University of Manchester Institute of Science and Technology

(UMIST)

PO Box 88, Manchester, M60 1QP, UK

Tel: 01612008930

Fax: 01612360409

Email: Simon.Hubbard@umist.ac.uk.

Location/Qualifiers

1. 701

/organism="Gallus gallus"

/mol\_type="mRNA"

/strain="Compton line 15r"

/db\_xref="taxon:9031"

/clone="CHEST480d7"

/sex="Female"

/tissue\_type="cerebellum"

/dev\_stage="adult"

/lab\_host="DH10B"

/clone\_lib="CSEQCHN69"

/note="Organ: brain; Vector: pluscript II KS(+); Site\_1:

Scor1; Site\_2: Not1; This normalized library was

constructed from 1 million independent clones. cDNA

synthesis was initiated using an oligo(dT) primer, using

methyated C in the first strand synthesis reaction.

Following this first strand reaction, double-stranded cDNA

was bluntend, ligated to NotI adapters, digested with

Scor1, size-selected, and cloned into the NotI and EcoRI

compatible sites of a custom modified MCS of the

pluscript (KS+) vector. The library was normalized in 2

rounds using conditions adapted from Soares et al., PNAS

(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6

(1996): 791, except that a significantly longer

reannealing hybridization was used."

reannealing hybridization was used."

reannealing hybridization was used."

reannealing hybridization was used."

reannealing hybridization was used."

reannealing hybridization was used."

reannealing hybridization was used."

reannealing hybridization was used."

reannealing hybridization was used."

reannealing hybridization was used."

reannealing hybridization was used."

reannealing hybridization was used."

reannealing hybridization was used."

reannealing hybridization was used."

reannealing hybridization was used."

reannealing hybridization was used."

reannealing hybridization was used."

reannealing hybridization was used."

reannealing hybridization was used."

reannealing hybridization was used."

reannealing hybridization was used."

reannealing hybridization was used."

reannealing hybridization was used."

reannealing hybridization was used."

reannealing hybridization was used."

reannealing hybridization was used."

reannealing hybridization was used."

reannealing hybridization was used."

reannealing hybridization was used."

reannealing hybridization was used."

reannealing hybridization was used."

reannealing hybridization was used."

reannealing hybridization was used."

reannealing hybridization was used."

reannealing hybridization was used."

reannealing hybridization was used."

reannealing hybridization was used."

reannealing hybridization was used."

reannealing hybridization was used."

reannealing hybridization was used."

Search completed: March 13, 2004, 09:16:27  
Job time : 2372 secs

Query Match 23.0%; Score 20; DB 13; Length 701;  
Best Local Similarity 100.0%; Pred.No. 4.4;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 13 TTAATTGAAGTGAATTTT 32  
|||||  
DB 128 TTAATTGAAGTGAATTTT 109



score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

OM nucleic - nucleic search, using sw model

Run on: March 12, 2004, 12:27:59 ; Search time 944.481 Seconds

(without alignments)  
13767.261 Million cell updates/sec

Title: US-09-966-880A-10\_COPY\_1\_300

Sequence: 1 ggggggcctgtaatcccaagct.....aatggcatggtgcaacctggc 300

Scoring table: IDENTITY\_NUC

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%

### Listing first 45 summaries

Database

1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_cm:\*  
5: gb\_ov:\*  
6: gb\_pac:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_scs:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_nm:\*  
20: em\_om:\*  
21: em\_ov:\*  
22: em\_ov:\*  
23: em\_pac:\*  
24: em\_pl:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_scs:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rod:\*  
36: em\_htg\_mad:\*  
37: em\_htg\_vrvt:\*  
38: em\_by:\*  
39: em\_hsgo\_hum:\*  
40: em\_hsgo\_mus:\*  
41: em\_hsgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

Result	No.	Score	Query Match	Length	DB	ID	Description
C	1	300	100.0	65664	6	BD016835	BD016835 Novel cyt
C	2	298.4	99.5	11304	6	BD016860	BD016860 Novel cyt
C	3	298.4	99.5	11304	9	AB040430	AB040430 Homo sapi
C	4	298.4	99.5	71132	9	AC092184	AC092184 Homo sapi
C	5	109.2	36.4	79528	6	BD176843	BD176843 Homo sapi
C	6	109.2	36.4	79528	9	HS466N1	297630 Human DNA s
C	7	107.8	35.9	42777	9	AL157391	AL157391 Human DNA
C	8	104.6	34.9	134060	9	AC010389	AC010389 Homo sapi
C	9	104.6	34.9	163673	9	AC093240	AC093240 Homo sapi
C	10	104	34.7	198521	2	AC145723	AC145723 Homo sapi
C	11	104	34.7	201337	2	AC141081	AC141081 Homo sapi
C	12	104	34.7	202689	2	AC034281	AC034281 Homo sapi
C	13	103.6	34.5	202533	2	AC144644	AC144644 Salmi'ri s
C	14	103.2	34.4	178016	2	AC146787	AC146787 Aotus nan
C	15	102.4	34.1	161864	9	AC007601	AC007601 Homo sapi
C	16	102.4	34.1	174136	2	AC126766	AC126766 Homo sapi
C	17	101.8	33.9	161090	6	AC117415	AC117415 Homo sapi
C	18	101.6	33.9	131329	9	AC004873	AC004873 Homo sapi
C	19	101.6	33.9	157335	9	HS50024	AL034380 Human DNA
C	20	101.6	33.9	170926	9	AC095034	AC095034 Homo sapi
C	21	101.6	33.9	182703	2	AC022939	AC022939 Homo sapi
C	22	101.6	33.9	185045	2	AC146466	AC146466 Callithr i
C	23	101.6	33.9	185784	2	AC009719	AC009719 Homo sapi
C	24	101.2	33.7	153311	9	AC011933	AC011933 Homo sapi
C	25	101.2	33.7	147534	9	AC068620	AC068620 Homo sapi
C	26	100.8	33.6	130934	9	EX005428	EX005428 Human DNA
C	27	100	33.3	97757	9	AC092402	AC092402 Homo sapi
C	28	100	33.3	118047	9	HS051399	AL094631 Human DNA
C	29	99.8	33.3	61450	9	HS15D7	AL031229 Human DNA
C	30	99.8	33.3	109745	9	AC025469	AC025469 Homo sapi
C	31	99.8	33.3	121394	2	AL137799	AL137799 Homo sapi
C	32	99.8	33.3	171710	2	AC068049	AC068049 Homo sapi
C	33	99.8	33.3	204151	2	AC145140	AC145140 Homo sapi
C	34	99.6	33.2	192653	9	AC124781	AC124781 Homo sapi
C	35	99.6	33.2	209775	2	AC027219	AC027219 Homo sapi
C	36	99.6	33.2	246322	2	AC133966	AC133966 Homo sapi
C	37	99.4	33.1	129957	9	HS415G2	Z83346 Human DNA s
C	38	99.2	33.1	161002	9	AL445664	AL445664 Human DNA
C	39	99.2	33.1	217022	2	AC146116	AC146116 Pan trogl
C	40	99	33.0	114450	9	AB023058	AB023058 Homo sapi
C	41	99	33.0	115133	9	AL669813	AL669813 Human DNA
C	42	99	33.0	119030	9	AL844851	AL844851 Human DNA
C	43	99	33.0	129818	9	AP000521	AP000521 Homo sapi
C	44	99	33.0	166842	9	HS1054A2	AL031651 Human DNA
C	45	99	33.0	319486	9	AF055066	AF055066 Homo sapi

## ALIGNMENTS

RESULT 1				
BD016835	BD016835	6564 bp	DNA	linear PAT 27-AUG-2002
LOCUS	BD016835			
DEFINITION	Novel cytidine deaminase.			
ACCESSION	BD016835			
VERSION	BD016835.1 GI:22558011			
KEYWORDS	JF 2001245669-A/8.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE	1 (bases 1 to 6564)			
AUTHORS	Honjo,T., and Muramatsu,M.			
TITLE	Novel cytidine deaminase			
JOURNAL	Patent: JP 2001245669-A 8 11-SEP-2001;			



	COMMENT	JAPAN TOBACCO INC./TASUKU HONDO OS Homo sapiens (human) PN JP 2001245669-A/8 PD 11-SEP-2001 PF 28-MAR-2000 JP 2000092981 PI TASUKU HONDO,MASAMICHI MORIMATSU PC C12N15/09,A6IK39/395,A6IK39/395,A6IP1/00,A6IP11/06,A6IP13/12, PC A6IP17/00, PC A6IP27/02,A6IP27/16,A6IP37/02,A6IP37/08,C07K16/18,C12N1/19, PC C12N1/21, PC C12N5/10,C12N9/78,C12P21/02,C12P21/08// (C12N1/21,C12R1:19), PC (C12N5/10,C12R1:91),C12N15/00,C12N5/00,(C12N5/00,C12R1:91) CC
	FEATURES	FEATURES SOURCE 1..6564 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606"
	ORIGIN	ORIGIN Query Match 100.0%; Score 300; DB 6; Length 6564; Best Local Similarity 100.0%; Pred. No. 3.2e-66; Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY		1 GGGGGCCGTATATCCCACTACTCTAGGAGGCGTGAGGCCGAGGAGMTCGCCGAGCCTGSCA 60
Db		1 GGGGGCCGTATATCCCACTACTCTAGGAGGCGTGAGGCCGAGGAGATCCCGAGCCTGSCA 60
QY		61 GATCGCCTTAGAGCCTGGAGAGTTGAGGCTACAGTAAAGCCAAGATCATGCAGTATACTTC 120
Db		61 GATCGCCTTAGAGCCTGGAGAGTTGAGGCTACAGTAAAGCCAAGATCATGCAGTATACTTC 120
QY		121 AGCCTGGCGCAAAAGTGAGACCGTACAAAAAAAATTAAAAAAGAATTATG 180
Db		121 AGCCTGGCGCAAAAGTGAGACCGTACAAAAAAAATTAAAAAAGAATTATG 180
QY		181 ATCAAGATCCCAACTGTATAAAAGTGGCCCTTAACAACAATTAAAGGTTTGAAGTTATTC 240
Db		181 ATCAAGATCCCAACTGTATAAAAGTGGCCCTTAACAACAATTAAAGGTTTGAAGTTATTC 240
QY		241 TGCAAGCGAAGAGAACCATCATCGAGGGGCTCTTCAGCATCGAGATGCGATGCGACCTGGT 300
Db		241 TGCAAGCGAAGAGAACCATCATCGAGGGGCTCTTCAGCATCGAGATGCGATGCGACCTGGT 300
	RESULT 2	
	LOCUS BDO16860 11204 bp DNA linear PAT 27-AUG-2002	
	DEFINITION Novel cytidine deaminase.	
	ACCSSION BDO16860	
	VERSION BDO16860.1 GI:22558036	
	KEYWORDS JP 2001245669-A/33.	
	SOURCE Homo sapiens (human)	
	ORGANISM Homo sapiens	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrates; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 11204) Honjo, T. and Muramatsu, M. Novel cytidine deaminase Patent: JP 2001245669-A 33 11-SEP-2001; JAFAN TOBACCO INC./TASUKU HONDO OS Homo sapiens (human) PN JP 2001245669-A/33 PD 11-SEP-2001 PF 28-MAR-2000 JP 2000092981 PI TASUKU HONDO,MASAMICHI MORIMATSU PC C12N15/09,A6IK39/395,A6IK39/395,A6IP1/00,A6IP11/06,A6IP13/12, PC A6IP17/00, PC A6IP27/02,A6IP27/16,A6IP37/02,A6IP37/08,C07K16/18,C12N1/19, PC C12N1/21, PC C12N5/10,C12N9/78,C12P21/02,C12P21/08// (C12N1/21,C12R1:19), PC (C12N5/10,C12R1:91),C12N15/00,C12N5/00,(C12N5/00,C12R1:91) CC	
	REFERENCE JOURNAL	
	AUTHORS	
	TITLE	
	COMMENT	
	Location/Qualifiers.	
	Key	

FEATURES

source

1. 11204

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

ORIGIN

Query Match 99.5%; Score 298.4; DB 6; Length 11204;

Best Local Similarity 99.7%; Pred. No. 8.3e-66;

Matches 299; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GGGGGCGCTGTATATCCAGCTACTCAGAGAGGCTGAGCAGAGAGATCCGCGAGCCTGCA 60

DB 5216 GGGGGCGCTGTATATCCAGCTACTCAGAGAGGCTGAGCAGAGAGATCCGCGAGCCTGCA 52755

OY 61 GATTCGCTTGAGCTGGAGAGTTGAGGCTACAGTAAGCCAAAGATCATGCTTACTTC 120

DB 5276 GATTCGCTTGAGCTGGAGAGTTGAGGCTACAGTAAGCCAAAGATCATGCTTACTTC 53358

OY 121 AGCTGGGCGGCAAGTGAGAGCCCTAACAAAAAATTTAAAAAAGAAATTAG 180

DB 5336 AGCTGGGCGGCAAGTGAGAGCCCTAACAAAAAATTTAAAAAAGAAATTAG 53955

OY 181 ATCAAGATCCACTGTATATAAAGTGCGCTTAACACCATTAAGAGTTGGAGTTATTC 240

DB 5396 ATCAAGATCCACTGTATATAAAGTGCGCTTAACACCATTAAGAGTTGGAGTTATTC 54555

OY 241 TGCAGGAGAAAGAGAACCATCATCGGGGGTCTTCAGCATGGGAATGGCATGGTGCACCTGCT 300

DB 5456 TGCAGGAGAAAGAGAACCATCATCGGGGGTCTTCAGCATGGGAATGGCATGGTGCACCTGCT 55155

RESULT 3

AB040430 11204 bp DNA linear PRI 03-OCT-2000

LOCUS

DEFINITION Homo sapiens A1d gene for activation-induced cytidine deaminase, complete cds.

ACCESSION AB040430

VERSION AB040430.1 GI:9988407

KEYWORDS A1d; activation-induced cytidine deaminase.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS Muto,T., Muramatsu,M., Taniwaki,M., Kinoshita,K. and Honjo,T.

TITLE 1 (sites)

JOURNAL JOURNAL MEDLINE 20408890

PUBMED 10950930

REFERENCE 2 (sites)

AUTHORS Rey,P., Muto,T., Levy,Y., Geisemann,F., Pieban,A., Sanal,O., Catalan,N., Forreille,M., Dutoourq-laglouise,R., Gennery,A., Tecchan,I., Essoy,F., Kayserili,H., Ugazio,A.G., Brousse,N., Muramatsu,M., Notarangelo,L.D., Kinoshita,K., Honjo,T., Fischer,A. and Durandy,A.

TITLE Activation-induced cytidine deaminase (A1d) deficiency causes the autosomal recessive form of the Hyper-Igm syndrome (HIGM2)

JOURNAL JOURNAL MEDLINE 20460541

PUBMED 11007475

REFERENCE 3 (bases 1 to 11204)

AUTHORS Muto,T., Muramatsu,M., Taniwaki,M., Kinoshita,K. and Honjo,T.

TITLE Direct Submission

JOURNAL Submitted (18-MAR-2000) Tasuku Honjo, Kyoto University, Department of Medical Chemistry, Faculty of Medicine, Yoshida, Sakyo-ku, Kyoto, Kyoto 606-8501, Japan (E-mail: honjo@omfou.med.kyoto-u.ac.jp, Tel:81-75-753-4371 (ex.4371), Fax:81-75-753-4388)

FEATURES

source

1. 11204

/organism="Homo sapiens"

/mol\_type="genomic DNA"







standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality/info/genbank.annotation.html>.

## QUALSTAT-REPORT

FEATURES  
source location/Qualifiers

1. 71132  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="12"  
/clone="RP11-438L7"  
/complement(1..1448)  
/note="overlaps bases 1..1448 of clone AC092490"  
/function="clone overlap"  
187..408  
/standard\_name="57233"  
439..560  
/standard\_name="92005"  
774..881  
/rpt\_family="MIR"  
903..1190  
/rpt\_family="AluSg"  
1191..1213  
/rpt\_family="AT\_rich"  
1591..1807  
/standard\_name="6612"  
1744..1819  
/standard\_name="8198"  
1966..2264  
/rpt\_family="AluSg"  
3296..3328  
/rpt\_family="(TTTC)n"  
4087..4249  
/rpt\_family="AluSg"  
4360..4654  
/rpt\_family="AluSg"  
complement(5140..5262)  
/rpt\_family="FLAM\_C"  
5275..5304  
/rpt\_family="AT\_rich"  
5305..5563  
/rpt\_family="AluDo"  
5564..5599  
/rpt\_family="AT\_rich"  
5644..5733  
/rpt\_family="MSTD"  
5734..6012  
/rpt\_family="AluSg"  
6020..6336  
/rpt\_family="AluSg"  
6337..6629  
/rpt\_family="MSTD"  
6630..6650  
/rpt\_family="(CA)n"  
complement(6711..6817)  
/rpt\_family="MIR"  
complement(7033..7327)  
/rpt\_family="AluSg"  
7607..7745  
/rpt\_family="FLAM\_A"  
7980..8047  
/rpt\_family="MIR"  
complement(8050..8549)

/rpt\_family="LTR47A"  
8567..8698  
/rpt\_family="MIR"  
complement(8815..9101)  
/rpt\_family="AluSg"  
complement(11227..12175)  
/rpt\_family="LTR5"  
12625..12900  
/standard\_name="87432"  
complement(13547..13664)  
/rpt\_family="L1MB5"  
complement(13665..13979)  
/rpt\_family="AluY"  
complement(13980..14193)  
/rpt\_family="L1MB5"  
14622..14924  
/rpt\_family="AluY"  
complement(15813..17136)  
/rpt\_family="L1PA8"  
complement(17272..17395)  
/rpt\_family="FLAM\_C"  
17426..17575  
/rpt\_family="MIR"

Query Match 99.5%; Score 298.4; DB 9; Length 71132;  
Best Local Similarity 99.7%; Pred. No. 9e-66; 1; Indels 0; Gaps 0;  
Matches 299; Conservative 0; Mismatches 1;

QY 1 GGGGGCTGTATTCAGCTACTCAGAGGCTGAGGAGGATCCGGAGCTGGCA 60  
DB 39760 GGGGGCTGTATTCAGCTACTCAGAGGCTGAGGAGGATCCGGAGCTGGCA 39819  
QY 61 GATTCGCTGAGCTGGAGGCTTACGCTACAGTAAAGCCAGTATTCAGTATTC 120  
DB 39820 GATTCGCTGAGCTGGAGGCTTACGCTACAGTAAAGCCAGTATTCAGTATTC 39879  
QY 121 AGCTGGGCGCAAGAGTGAAGCGTAAACAAAAAATTTAAAGAAATTTAG 180  
DB 39880 AGCTGGGCGCAAGAGTGAAGCGTAAACAAAAAATTTAAAGAAATTTAG 39939  
QY 181 ATCAAGATCCAACTGTAAAAAGTGGCTTAAACACCATTTAAAGTTGAGTTATTC 240  
DB 39940 ATCAAGATCCAACTGTAAAAAGTGGCTTAAACACCATTTAAAGTTGAGTTATTC 39999  
QY 241 TGCAGGCAAGAAACATCAGGAGGCTTTCAGATGAGAAAGGCTGATGACCTGCT 300  
DB 40000 TGCAGGCAAGAAACATCAGGAGGCTTTCAGATGAGAAAGGCTGATGACCTGCT 40059

RESULT 5  
LOCUS BD176843 79528 bp DNA linear PAT 18-MAR-2003  
DEFINITION A method of predicting cancer condition.  
ACCESSION BD176843  
VERSION BD176843.1 GI:29122555  
KEYWORDS WO 02072828-A/6.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 79528)  
Kato, K., Iwao, K., Noguchi, S. and Matoba, R.  
A method of predicting cancer condition.  
Patent: WO 02072828-A 6 19-SEP-2002;  
DNA CHIP RESEARCH INC. HITACHI SOFTWARE ENGINEERING CO LTD, KIKUYA  
KATO, KYOKO IMAO, SHINZABURO NOGUCHI, RIO MATOBA  
OS Homo sapiens (human)  
PN WO 02072828-A/6  
PD 19-SEP-2002  
PF 07-MAR-2002 WO 2002JP002153  
PR 14-MAR-2001 JP 01P 073063, 06-APR-2001 JP 01P 108503 PR  
P2-AUG-2001 JP 01P 234807  
PI KIKUYA KATO, KYOKO IMAO, SHINZABURO NOGUCHI, RIO MATOBA PC



```

FEATURES
  source
    location/Qualifiers
      1..79528
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"

ORIGIN
Query Match      36.4%; Score 109.2; DB 6; Length 79528;
Best Local Similarity 72.7%; Pred. No. 1,9e-17;
Matches 141; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 1 GGGGGCTGTATATCCAGTACTACAGAGGCTGAGCGAGAGGANTCGCGAGCCTTGACA 60
Db 64094 GGGGGCTGTATATCCAGTACTACAGAGGCTGAGCGAGAGGANTCGCGAGCCTTGACA 64035
QY 61 GATCTGCTTAGACCTGAGGAGTTGAGGCTACAGTAAAGCCAAATCATGCCAGTATCTTC 120
Db 64034 GAATCGCTTAGACCTGAGGAGGAGTTGAGGAGCCAAATGTGCGCATCTGACTTC 63975
QY 121 AGCTGCGGCCACAAAGTGAAGACCTTACAAAATTTTAAATTTAAATTTAAATTTG 180
Db 63974 AGCTGCGAGACAAAGTGAAGACCTTACAAAATTTTAAATTTTAAATTTAAATTA 63915
QY 181 ATCAAGATCCACT 194
Db 63914 ATTAATTAATTAAT 63901

RESULT 6
H5466N1/c
LOCUS
DEFINITION
H5466N1 79528 bp DNA linear PRI 05-JUN-2003
Human DNA sequence from clone R3-46N1 on chromosome 22q12-13
Contains the H1FO gene for H1 histone family member 0, the GCAT
gene for glycine C-acetyltransferase (2-amino-3-ketobutyrate
coenzyme A ligase), the GALR3 gene for galactin receptor, the gene
for a novel protein similar to ANK3 (ankyrin 3, node of Ranvier
(ankyrin G)), the 5' end of the gene for proteins HSPC025 and
HSPC021 (similar to C. elegans PAT-3 alcohol dehydrogenase), ESTs,
STS, GSSs and eight putative CpG islands, complete sequence.
297630
297630.11 GI:4582128
HNG: alcohol dehydrogenase; ANK3: ankyrin; Cpg island: PAT-3;
galactin receptor; GALR3; GCAT; H1FO; histone; HSPC021; HSPC025.
SOURCE
Organism
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 79528)
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Direct Submission
Submitted (05-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
hinxton@sanger.ac.uk; Clone requests: clonerequest@sanger.ac.uk
On Apr 13, 1999 this sequence version replaced GI:4581359.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
-----
Genome Center
Centre: Wellcome Trust Sanger Institute
Center code: SC

```

Web site: <http://www.sanger.ac.uk>  
Contact: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr22>

RP3-166N1 is from the library RPCI-3 constructed by the group of Plier de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pCYPAC2

IMPORTANT: This sequence is not the entire insert of clone RP3-166N1. It may be shorter because we sequence overlapping sections only once, except for a short overlap.

The true left end of clone RP5-101d13 is at 79429 in this sequence. The true right end of clone RP1-37d16 is at 100 in this sequence.

FEATURES	
source	Location/Qualifiers
	1..79528
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="22"
	/map="q12-13"
	/clone="RP3-466N1"
	/clone_1kb="RPCT-3"
repeat_region	8..17
	/note="2.0 copies 5 mer AGCTC 20% conserved"
	39..174
repeat_region	/note="AluSg/x repeat: matches 1..136 of consensus"
repeat_region	175..492
	/note="AluSx repeat: matches 1..312 of consensus"
repeat_region	494..505
	/note="3.0 copies 4 mer GAA 24% conserved"
repeat_region	850..1133
	/note="AluJo repeat: matches 22..312 of consensus"
repeat_region	1134..1446
	/note="AluSg repeat: matches 1..313 of consensus"
repeat_region	1514..1597
	/note="MR repeat: matches 180..262 of consensus"
repeat_region	1837..2110
	/note="AluJo repeat: matches 13..294 of consensus"
repeat_region	2113..2253
	/note="FRLM_C repeat: matches 1..133 of consensus"
repeat_region	2281..2280
	/note="3.3 copies 6 mer AACAC 26% conserved"
repeat_region	2329..2319
	/note="2.3 copies 9 mer TTGATCAG 33% conserved"
repeat_region	2537..2555
	/note="13.8 copies 5 mer TTGT 38% conserved"
repeat_region	complement(2556..2759)
	/note="AluSx repeat: matches 146..311 of consensus"
repeat_region	complement(2760..3070)
	/note="AluYc2 repeat: matches 1..310 of consensus"
repeat_region	complement(3071..3302)
	/note="AluSx repeat: matches 4..146 of consensus"
repeat_region	3305..3215
	/note="11.0 copies 1 mer T 22% conserved"
repeat_region	complement(3320..3332)
	/note="AluSx repeat: matches 1..312 of consensus"
repeat_region	3555..3766
	/note="12 repeat: matches 2443..2691 of consensus"
repeat_region	3842..3908
	/note="L1MG repeat: matches 5672..5727 of consensus"
repeat_region	3909..4040



```

/note="FLAM_C repeat: matches 3. .133 of consensus"
4041. .4072
repeat_region
/note="LIME repeat: matches 5727. .5761 of consensus"
complement(4073. .4399)
repeat_region
/note="MLRID repeat: matches 162. .505 of consensus"
4400. .4697
repeat_region
/note="AluSp repeat: matches 1. .299 of consensus"
complement(4698. .4860)
repeat_region
/note="MLRID repeat: matches 4. .162 of consensus"
4861. .4886
repeat_region
/note="2.2 copies 12 mer CACACAGTACC 52% conserved"
4897. .4911
repeat_region
/note="2.5 copies 6 mer GGGCTG 30% conserved"
5073. .5113
repeat_region
/note="L1MA1 repeat: matches 6214. .6256 of consensus"
5114. .5420
repeat_region
/note="AluSx repeat: matches 1. .307 of consensus"
5421. .5473
repeat_region
/note="L1MA1 repeat: matches 6256. .6302 of consensus"
5476. .5498
repeat_region
/note="2.3 copies 1 mer A 28% conserved"
complement(5503. .5977)
repeat_region
/note="L2 repeat: matches 3136. .3220 of consensus"
6050. .6393
repeat_region
/note="MER83 repeat: matches 1. .355 of consensus"
complement(6394. .6705)
repeat_region
/note="AluY repeat: matches 1. .311 of consensus"
complement(6706. .7009)
repeat_region
/note="AluSx repeat: matches 1. .305 of consensus"
7010. .7126
repeat_region
/note="MER83 repeat: matches 353. .448 of consensus"
complement(7145. .7445)
repeat_region
/note="L1MA4 repeat: matches 5936. .6226 of consensus"
7446. .7455
repeat_region
/note="10.0 copies 1 mer T 20% conserved"
complement(7466. .7588)
repeat_region
/note="AluY repeat: matches 1. .123 of consensus"
complement(7589. .7887)
repeat_region
/note="AluSp repeat: matches 1. .302 of consensus"
complement(7898. .8117)
repeat_region
/note="AluSg/X repeat: matches 83. .300 of consensus"
complement(8121. .8453)
repeat_region
/note="L1MA4 repeat: matches 5618. .5949 of consensus"
8456. .8510
repeat_region
/note="AluSg/X repeat: matches 119. .133 of consensus"
8511. .8807
repeat_region
/note="AluSx repeat: matches 1. .298 of consensus"
8808. .8983
repeat_region
/note="AluSg/X repeat: matches 133. .306 of consensus"
8995. .9004
repeat_region
/note="3.3 copies 3 mer ACC 20% conserved"
9031. .9082
repeat_region
/note="L2 repeat: matches 3106. .3160 of consensus"
complement(9083. .9386)
repeat_region
/note="AluSg repeat: matches 1. .304 of consensus"
9387. .9522
repeat_region
/note="L2 repeat: matches 3160. .3282 of consensus"
9546. .9852
repeat_region
/note="AluSg1 repeat: matches 1. .306 of consensus"
9947. .9960
repeat_region
/note="2.3 copies 6 mer GGGGTG 28% conserved"
complement(10080. .10164)
repeat_region
/note="MIR repeat: matches 109. .206 of consensus"
10110. .10367
repeat_region
/note="match: GSS: Em:AQ416488"
10157. .10649
misc_feature
/evidence=not_experimental
10299. .10316
repeat_region
/note="2.2 copies 8 mer CCGCCGTT 27% conserved"
10478. .10512
repeat_region
/note="7.0 copies 5 mer GAGGC 61% conserved"
10777. .10788
repeat_region

```

```

/note="2.4 copies 5 mer AAATG 24% conserved"
11262. .11272
repeat_region
/note="2.2 copies 5 mer CTTC 22% conserved"
11328. .11341
repeat_region
/note="2.3 copies 6 mer CTCCTC 28% conserved"
11490. .11499
repeat_region
/note="2.0 copies 5 mer CCCCT 20% conserved"
11583. .11595
repeat_region
/note="2.2 copies 6 mer ACTCC 26% conserved"
12535. .13553
misc_feature
/note="CpG island"
/evidence=not_experimental
12598. .12612
repeat_region
/note="3.0 copies 5 mer CGCT 21% conserved"
12919. .12929
repeat_region
/note="3.7 copies 3 mer GGC 22% conserved"

```

```

Query Match 36.4%; Score 109.2; DB 9; Length 79528;
Best Local Similarity 72.7%; Pred. No. 1.9e-17;
Matches 141; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

```

```

QY 1 GGGGGCCCTGATCCAGCTACTCAGAGGCTGAGAGGAGATCCGGAGCTTGCA 60
DB 64094 GGGGCGCTGTATCCAGCTACGCGAGGCTGAGAGGATTCCTGAACCTAGGA 64035
QY 61 GATCGCTGAGCTGGAGGCTTACGCTACAGTAAGCAAGATCATGCCAGTACTTC 120
DB 64034 GATGCTGTGAACCTGGAGGCGAGGTTGCGAGACCAAGATTGCCACTGTTCTC 63975
QY 121 AGCCTGGCGCAAGAGTGAGCCGTAAACAAAAAATTTAAAGAAATTTAG 180
DB 63974 AGCCTGGAGAGCAGACGAGCTCATCTCAAAAAATATATATATATATATTA 63915
QY 181 ATCAAGATCCACT 194
DB 63914 ATTAATTAATTAAT 63901

```

```

RESULT 7
AL157391/c 47177 bp DNA linear PRI 10-AUG-2001
LOCUS Human DNA sequence from clone RP11-271M1 on chromosome 10, complete
DEFINITION
ACCESSION AL157391
VERSION AL157391.11 GI:15149560
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 47177)
REFERENCE
AUTHORS Chapman,J.
TITLE Direct Submission
JOURNAL Submitted (10-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Aug 13, 2001 this sequence version replaced gi:15021290.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em, EMBL; Sw,
SWISSPROT; Tr, TREMBL; Wp, WORMBASE; information on the WORMBASE

```



database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Chr10>  
RP11-271M1 is from the library RPC1-11.1 constructed by the group of Pieter de Jong. For further details see  
<http://www.chori.org/bacpac/home.htm>  
VECTOR: pBAC3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-271M1 it may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.  
The true left end of clone RP11-271M1 is at 1 in this sequence. The true left end of clone RP11-455B2 is at 45178 in this sequence.

## FEATURES

Source  
1..47177  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="10"  
/clone="RP11-271M1"  
/clone\_id="RPC1-11.1"  
1..2296  
/note="HERVE repeat: matches 1108..3423 of consensus"  
2313..2628  
repeat\_region  
/note="4 copies 79 mer 67% conserved"  
2726..2883  
repeat\_region  
/note="2 copies 79 mer 82% conserved"  
2797..3100  
repeat\_region  
/note="4 copies 76 mer 70% conserved"  
3348..3581  
repeat\_region  
/note="3 copies 78 mer 79% conserved"  
3371..3895  
misc\_feature  
/note="Cpg Island"  
/evidence="not\_experimental"  
3499..3806  
repeat\_region  
/note="4 copies 77 mer 75% conserved"  
3616..3843  
repeat\_region  
/note="3 copies 76 mer 78% conserved"  
3734..3889  
repeat\_region  
/note="2 copies 78 mer 82% conserved"  
3949..4134  
repeat\_region  
/note="3 copies 62 mer 75% conserved"  
4101..4352  
repeat\_region  
/note="4 copies 63 mer 77% conserved"  
4290..4413  
repeat\_region  
/note="2 copies 62 mer 83% conserved"  
4525..4902  
repeat\_region  
/note="6 copies 63 mer 77% conserved"  
4813..5122  
repeat\_region  
/note="5 copies 62 mer 68% conserved"  
4947..5072  
repeat\_region  
/note="2 copies 63 mer 85% conserved"  
5164..5415  
repeat\_region  
/note="4 copies 63 mer 75% conserved"  
5501..5815  
repeat\_region  
/note="5 copies 63 mer 75% conserved"  
5772..6005  
repeat\_region  
/note="3 copies 78 mer 85% conserved"  
6094..6245  
repeat\_region  
/note="2 copies 76 mer 85% conserved"  
6294..6445  
repeat\_region  
/note="2 copies 76 mer 86% conserved"  
6317..6457  
misc\_feature  
/note="match: STS: Em:HSFH09B8"  
8157..8457  
repeat\_region  
/note="AluSg repeat: matches 1..299 of consensus"  
8524..8636  
repeat\_region  
/note="12 repeat: matches 2596..2710 of consensus"  
9039..9348  
repeat\_region  
/note="AluSg repeat: matches 1..308 of consensus"  
9743..9955

repeat\_region  
/note="MIR repeat: matches 32..252 of consensus"  
10799..11045  
repeat\_region  
/note="11MB3 repeat: matches 5896..6151 of consensus"  
11046..11335  
repeat\_region  
/note="AluSx repeat: matches 12..304 of consensus"  
11336..11461  
repeat\_region  
/note="11MB3 repeat: matches 5766..5896 of consensus"  
11462..11771  
repeat\_region  
/note="AluSx repeat: matches 6..312 of consensus"  
11772..12766  
repeat\_region  
/note="11MB3 repeat: matches 4818..5766 of consensus"  
12812..13119  
repeat\_region  
/note="AluSg repeat: matches 1..308 of consensus"  
13200..13406  
repeat\_region  
/note="11MB4 repeat: matches 5961..6130 of consensus"  
13407..13686  
repeat\_region  
/note="AluSx repeat: matches 1..280 of consensus"  
13687..13865  
repeat\_region  
/note="11MB4 repeat: matches 5771..5961 of consensus"  
14056..14388  
repeat\_region  
/note="AluSg repeat: matches 1..312 of consensus"  
14547..14692  
repeat\_region  
/note="MER5B repeat: matches 25..172 of consensus"  
14798..14995  
repeat\_region  
/note="MER6B repeat: matches 5..210 of consensus"  
complement(15721..16199)  
/note="match: STS: Em:G55345"  
16895..17117  
repeat\_region  
/note="AluSg repeat: matches 1..222 of consensus"  
17114..17365  
repeat\_region  
/note="4 copies 63 mer 86% conserved"  
17293..17444  
repeat\_region  
/note="2 copies 76 mer 83% conserved"  
17389..17514  
repeat\_region  
/note="2 copies 63 mer 83% conserved"  
17446..17597  
repeat\_region  
/note="2 copies 76 mer 87% conserved"  
17528..17716  
repeat\_region  
/note="3 copies 63 mer 89% conserved"  
17645..17796  
repeat\_region  
/note="2 copies 76 mer 89% conserved"  
17721..17972  
repeat\_region  
/note="4 copies 63 mer 88% conserved"  
17913..18444  
repeat\_region  
/note="7 copies 76 mer 82% conserved"  
19048..19883  
repeat\_region  
/note="11 copies 76 mer 63% conserved"  
19355..19654  
repeat\_region  
/note="2 copies 150 mer 86% conserved"  
19645..19800  
repeat\_region  
/note="2 copies 78 mer 82% conserved"  
20048..20351  
repeat\_region  
/note="4 copies 76 mer 79% conserved"  
20414..20641  
repeat\_region  
/note="3 copies 76 mer 71% conserved"  
20697..22765  
repeat\_region  
/note="HERVE repeat: matches 3122..5178 of consensus"  
23280..23589  
repeat\_region  
/note="AluSx repeat: matches 1..309 of consensus"  
23984..24293  
repeat\_region  
/note="AluSx repeat: matches 2..311 of consensus"  
24761..25049  
repeat\_region  
/note="AluSx repeat: matches 4..296 of consensus"  
25132..25437  
repeat\_region  
/note="AluSx repeat: matches 1..307 of consensus"  
25926..26093  
repeat\_region  
/note="AluSg/x repeat: matches 143..309 of consensus"  
26150..26300  
repeat\_region  
/note="AluY repeat: matches 135..302 of consensus"  
26303..26691  
repeat\_region  
/note="11MB6 repeat: matches 5700..6095 of consensus"  
26722..27017  
repeat\_region  
/note="AluSx repeat: matches 1..298 of consensus"



Query Match	35.9%	Score 107.8;	DB 9;	Length 47177;
Best Local Similarity	76.0%	Pred. No. 4.3e-17;		
Matches 133; Conservative	0;	Mismatches 42;	Indels 0;	Gaps

RESULT 8	
AC010389/c	
LOCUS	AC010389
DEFINITION	Homo sapiens chromosome 5 clone CTD-207715, complete sequence.
ACCESSION	AC010389
VERSION	AC010389.8 GI:16195201
	134060 bp DNA linear PRI 17-OCT-2001

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE	1 (bases 1 to 134060)
AUTHORS	DOE Joint Genome Institute and Stanford Human Genome Center
TITLE	Direct Submission

Unpublished  
(bases 1 to 134060)  
DOE Joint Genome Institute

**TITLE** Direct Submission  
**JOURNAL** Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE	3 (bases 1 to 134060)
AUTHORS	DOE Joint Genome Institute and Stanford Human Genome Center
TITLE	Direct Submission

REFERENCE  
JOURNAL  
Submitted (30-JUN-2000) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA  
4 (bases 1 to 134060)

**AUTHORS** DOE Joint Genome Institute and Stanford Human Genome Center.  
**TITLES** Direct Submission  
**JOURNAL** Submitted (17-OCT-2001) DOE Joint Genome Institute, 2800 Mitchell

COMMENT  
Drive, Walnut Creek, CA 94598, USA  
On Oct 17, 2001 this sequence version replaced gi:8844109

www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center

```

www-shgc.stanford.edu
Quality: Phrap Quality >=40 99.9% of Sequence
Estimated Total Number of Errors is 0.2.
Location/Qualifiers
1. .134060
/covariate=134060.conf

```

Query Match	34.9%	Score 104.6;	DB 9;	Length 134060;
Best Local Similarity	73.2%	Pred. No. 3e-16;		
Matches 134; Conservative	0;	Mismatches 49;	Indels 0;	Gaps 0;

0y 2GGGGGCTGTAAATCCAGCTACTCTCAGGAGGCTGAGCAGAGGATCCGGAGGCTGGCAG 61  
 Db 57421 GGGGCTGTAAATCCAGCTACTCTCAGGAGGCTGAGCAGAGGATCCCTTGAACGACAGAG 57366

62 ATCTGCTGAGCCTGGAGTTGAGGCTACAGTAAGCCAGATCATGCCAGTAACTTCA 121

DB 57361AAATCACTTGAACCGAGGATGGAGGCTTGCACTGAGGCTAAAGATTGGCCACATCGACATCCCA 5/30

QY 122 GCTTGGGCGCACAAGTAGACCCGTAACAAAAAATTTAAAAAAGAAATTTGCA 181

Db 57301 GTCGGGCTCAGAGTGAAGCTCCATCTCAAAAAAAAAAAAAAAAAAAGAGCTTGT 5724

QY	182	TCA	184
Db	57241	TAA	57239

RESULT 9  
AC093240/c

LOCUS	AC093240	163673 bp	DNA	linear	PRI 31-JAN-93
DEFINITION	Homo sapiens chromosome 5 clone RP11-103P4, complete sequence				
ACCESSION	AC093240				

VERSION	AC093240.2	GI:18449986
KEYWORDS	HTG.	
SOURCE	Homo sapiens (human)	

ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
----------	---

REFERENCE 1 (bases 1 to 163673)  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center  
TITLE Direct Submission

Unpublished  
2 (bases 1 to 163673)  
DOE Joint Genome Institute

**TITLE** Direct Submission  
**Submitted** (16-AUG-2001) **Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA**

REFERENCE	3 (bases 1 to 163673)
AUTHORS	DOE Joint Genome Institute and Stanford Human Genome Center
TITLE	Direct Submission

**JOURNAL** Submitted (31-JAN-2002) DOE Joint Genome Institute, 2800 Mi  
Drive, Walnut Creek, CA 94598, USA  
**COMMENT** On Jan 31, 2002 this sequence version replaced gi:15193374

Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center

www-shgc.stanford.edu  
Quality: Phrap Quality >=40 99.9% of Sequence  
Estimated Total Number of Errors is 0.2.

FEATURES  
source

```
/organism="Homo sapiens"  
/mol_type="genomic DNA"  
/db_xref="taxon:9606"  
/chromosome="5"  
/clone="RP11-103P4"
```

## ORIGIN







```
* 76134 79417: contig of 3284 bp in length
* 79418 79517: gap of unknown length
* 79518 81816: contig of 2299 bp in length
* 81817 81916: gap of unknown length
* 81917 84825: contig of 2909 bp in length
* 84826 84925: gap of unknown length
* 84926 87053: contig of 2128 bp in length
* 87054 87153: gap of unknown length
* 87154 90407: contig of 3254 bp in length
* 90408 90507: gap of unknown length
* 90508 94550: contig of 4043 bp in length
* 94551 94651: gap of unknown length
* 94651 96993: contig of 2343 bp in length
* 96994 97094: gap of unknown length
* 97094 101223: contig of 4130 bp in length
* 101224 101323: gap of unknown length
* 101324 104392: contig of 3069 bp in length
* 104393 104492: gap of unknown length
* 104493 108893: contig of 4401 bp in length
* 108894 108993: gap of unknown length
* 108994 113350: contig of 4357 bp in length
* 113351 113450: gap of unknown length
* 113451 118703: contig of 5253 bp in length
* 118704 118803: gap of unknown length
* 118804 121418: contig of 2615 bp in length
* 121419 121518: gap of unknown length
* 121519 126615: contig of 5097 bp in length
* 126616 126715: gap of unknown length
* 126716 131029: contig of 4314 bp in length
* 131030 131129: gap of unknown length
* 131130 137699: contig of 6570 bp in length
* 137700 137799: gap of unknown length
* 137800 145730: contig of 5931 bp in length
* 143731 143830: gap of unknown length
* 143831 151122: contig of 7292 bp in length
* 151123 151223: gap of unknown length
* 151224 159620: contig of 8398 bp in length
* 159621 159720: gap of unknown length
* 159721 172370: contig of 12550 bp in length
* 172371 198521: contig of 26151 bp in length.
* 172371 198521: contig of 26151 bp in length.

FEATURES
source      1..198521
             /organism="Homo sapiens"
             /mol_type="genomic DNA"
             /db_xref="taxon:9606"
             /chromosome="19"
             /clone="XXfos-8909F12"

ORIGIN
Query Match      34.7%; Score 104; DB 2; Length 198521;
Best Local Similarity 72.8%; Pred. No. 4.3e-16;
Matches 134; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

1 GGGGGCTGTATCCAGCTACTCAGAGGCTGAGGCGAGAGATCCGCGAGCTTGCA 60
13253 GGGTGCCTGTATCCAGCTACTCAGAGGCTGAGGCGAGAGATCGCTTGAGGCGAGA 13194

61 GATTCGCTGAGCCTGGAGGTTGAGGCTACAGTAAGCCAGATCCAGTATCTTC 120
DB 13193 GAATCGCTTGAAACCGAGAGCGGAGGTTGCAATGACCCAGATCGTCCACTGCCTCC 13134

121 AGCTGGGCGCAAGTGAAGCCGTAAACAAAAAATTTAAAAAAGAAATTTAG 180
DB 13133 AGCTGGGCGCAAGTGAAGCTATCTCAATAAATAACAGATTAAAGTAATAAAAA 13074

181 ATCA 184
DB 13073 ATTA 13070
```

```
LOCUS AC141081 201237 bp DNA linear HTG 07-MAR-2003
DEFINITION Homo sapiens chromosome 16 clone RP11-927N1, WORKING DRAFT
SEQUENCE 3 unordered pieces.
AC141081
ACCESSION AC141081.1 GI:26875941
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 201237)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 201237)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (07-MAR-2003) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT -----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 16S0785
Center clone name: RP11-927N1
-----
Summary Statistics
Consensus quality: 200303 bases at least Q40
Consensus quality: 200551 bases at least Q30
Consensus quality: 200605 bases at least Q20
Estimated insert size: 175000; agarose-fp estimation
Estimated insert size: 201037; sum-of-contigs estimation
Quality coverage: 12.74 in Q20 bases; agarose-fp estimation
Quality coverage: 11.09 in Q20 bases; sum-of-contigs estimation.
NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 34897: contig of 34897 bp in length
* 34898 34997: gap of unknown length
* 34998 88292: contig of 53295 bp in length
* 88293 88392: gap of unknown length
* 88393 201237: contig of 112845 bp in length.

FEATURES
source      1..201237
             /organism="Homo sapiens"
             /mol_type="genomic DNA"
             /db_xref="taxon:9606"
             /chromosome="16"
             /clone="RP11-927N1"
             /clone_id="RP11 human BAC library 11"

ORIGIN
Query Match      34.7%; Score 104; DB 2; Length 201237;
Best Local Similarity 72.8%; Pred. No. 4.3e-16;
Matches 134; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

1 GGGGGCTGTATCCAGCTACTCAGAGGCTGAGGCGAGAGATCCGCGAGCTTGCA 60
DB 144542 GGGTGCCTGTATCCAGCTACTCAGAGGCTGAGGCGAGAGATCGCTTGAGGCGAGA 144601

61 GATTCGCTGAGCCTGGAGGTTGAGGCTACAGTAAGCCAGATCCAGTATCTTC 120
DB 144602 GAATCGCTTGAAACCGAGAGCGGAGGTTGCAATGACCCAGATCGTCCACTGCCTCC 144661

121 AGCTGGGCGCAAGTGAAGCCGTAAACAAAAAATTTAAAAAAGAAATTTAG 180
```



Db 144662 AGCCTGGCAGACAGAGTGAAGCTCTATCTCAATAAAGAGATTAAAGTAATMAA 144721  
ATCA 184  
Db 144722 ATTA 144725

RESULT 12  
AC034281  
LOCUS  
DEFINITION  
Homo sapiens chromosome 16 clone RP11-65606 map 16, WORKING DRAFT  
SEQUENCE, 28 unordered pieces.  
AC034281  
AC034281.3 GI:10047874  
KEYWORDS  
HTG; HTGS PHASE1; HTGS\_DRAFT.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
1 (bases 1 to 202689)  
Birren, B., Linton, J., Nusbaum, C. and Lander, E.  
Homo sapiens chromosome 16, clone RP11-65606  
Unpublished  
2 (bases 1 to 202689)  
Birren, B., Linton, J., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,  
Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G.,  
Campopiano, A., Castle, A., Choepel, Y., Collangelo, M., Collins, S.,  
Collins, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,  
Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,  
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,  
Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L.,  
Howard, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,  
Klein, J., Larocque, K., Lamazares, R., Landers, T., Lehoczy, J.,  
Levine, R., Liu, C., Locke, K., MacDonald, P., Marguis, N.,  
McCarthy, M., McGowan, P., McGurk, A., McKernan, K., McPherson, R.,  
Meldrum, J., Menus, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J.,  
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,  
O'Neill, D., O'Neil, T.M., Oliver, J., Peterson, K., Pierre, N.,  
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
Strand-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
Tessfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigglio, J.,  
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.U.,  
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE  
Direct Submission  
Submitted (05-APR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 202689)  
Birren, B., Linton, J., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
Anderson, S., Barna, N., Bastien, V., Beda, F., Boguslavsky, L.,  
Boukhalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A.,  
Choepel, Y., Collangelo, M., Collins, S., Collangelo, A., Cooke, P.,  
Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Ferreira, P.,  
Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M.,  
Graham, L., Grand-Pierre, N., Hagos, B., Heatford, A., Horton, L.,  
Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Larocque, K.,  
Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, C., Liu, G.,  
Macdonald, P., Marguis, N., McCarthy, M., McGowan, P., McKernan, K.,  
McPherson, R., Meldrum, J., Menus, L., Mihova, T., Mienga, V.,  
Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T.,  
O'Donnell, P., O'Neil, D., O'Neil, T.M., Oliver, J., Peterson, K.,  
Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riback, M., Riley, R.,  
Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P.,  
Sougnuez, C., Spencer, B., Strand-Thomann, N., Stojanovic, N.,  
Strasas, N., Subramanian, A., Talamas, J., Tessfaye, S., Theodore, J.,  
Tirrell, A., Travers, M., Trigglio, J., Vassiliev, H., Viel, R., Vo, A.,  
Wilson, B., Wu, X., Wyman, D., Ye, W.U., Young, G., Zainoun, J.,  
Zimmer, A. and Zody, M.

TITLE  
Direct Submission  
Submitted (24-APR-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Sep 10, 2000 this sequence version replaced gi:7770529.

All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
Project Information  
Center project name: 18279  
Center clone name: 656\_O\_6  
Sequencing Summary Statistics  
Vector: M13; M77815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 182566 bases at least Q40  
Consensus quality: 192896 bases at least Q30  
Consensus quality: 196996 bases at least Q20  
Insert size: 210000; agarose-fp  
Insert size: 199989; sum-of-contigs  
Quality coverage: 3.9 in Q20 bases; agarose-fp  
Quality coverage: 4.1 in Q20 bases; sum-of-contigs  
\*\*\*\*\*  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 28 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
1  
216: contig of 216 bp in length  
217  
316: gap of 100 bp  
317  
1822: contig of 1506 bp in length  
1823  
1922: gap of 100 bp  
1923  
2968: contig of 1046 bp in length  
2969  
3068: gap of 100 bp  
3069  
4380: contig of 1312 bp in length  
4381  
4480: gap of 100 bp  
4481  
6135: contig of 1655 bp in length  
6136  
6235: gap of 100 bp  
6236  
9213: contig of 2978 bp in length  
9214  
9313: gap of 100 bp  
9314  
11857: contig of 2544 bp in length  
11858  
11957: gap of 100 bp  
11958  
14517: contig of 2560 bp in length  
14518  
14617: gap of 100 bp  
14618  
32144: contig of 17527 bp in length  
32145  
32244: gap of 100 bp  
32245  
35656: contig of 3112 bp in length  
35657  
35756: gap of 100 bp  
35757  
40409: contig of 4653 bp in length  
40410  
40509: gap of 100 bp  
40510  
46096: contig of 5887 bp in length  
46097  
46196: gap of 100 bp  
46197  
50343: contig of 4147 bp in length  
50344  
50443: gap of 100 bp  
50444  
54901: contig of 4458 bp in length  
54902  
55001: gap of 100 bp  
55002  
59436: contig of 4435 bp in length  
59437  
59536: gap of 100 bp  
59537  
66852: contig of 7116 bp in length  
66853  
69537: gap of 100 bp  
69538  
75809: contig of 8857 bp in length  
75810  
75909: gap of 100 bp  
75910  
85003: contig of 9100 bp in length  
85004  
85109: gap of 100 bp  
85110  
94166: contig of 9057 bp in length  
94167  
94266: gap of 100 bp  
94267  
104445: contig of 10179 bp in length  
104446  
104545: gap of 100 bp  
104546  
114898: contig of 10353 bp in length  
114899  
114998: gap of 100 bp



114999 126032: contig of 11034 bp in length  
\* 126033 126132: gap of 100 bp  
\* 126133 139665: contig of 13533 bp in length  
\* 139666 139765: gap of 100 bp  
\* 139766 151421: contig of 11656 bp in length  
\* 151422 151521: gap of 100 bp  
\* 151522 169514: contig of 17993 bp in length  
\* 169515 169614: gap of 100 bp  
\* 169615 183307: contig of 13693 bp in length  
\* 183308 183407: gap of 100 bp  
\* 200849: contig of 17442 bp in length  
\* 183408 200850: gap of 100 bp  
\* 200850 202689: contig of 1740 bp in length.  
\* 200950

FEATURES  
SOURCE Location/Qualifiers

1..202689  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/map="16"  
/clone="RP11-656C6"  
/clone\_id="RP11 Human Male BAC"  
misc\_feature 1..216  
/note="assembly\_fragment  
clone end:SP6  
vector\_side:left"  
misc\_feature 317..1822  
/note="assembly\_fragment"  
misc\_feature 1923..2968  
/note="assembly\_fragment"  
misc\_feature 3069..4380  
/note="assembly\_fragment"  
misc\_feature 4481..6135  
/note="assembly\_fragment"  
misc\_feature 6236..9213  
/note="assembly\_fragment"  
misc\_feature 9314..11857  
/note="assembly\_fragment"  
misc\_feature 11958..14517  
/note="assembly\_fragment"  
misc\_feature 14618..32144  
/note="assembly\_fragment"  
misc\_feature 32245..35656  
/note="assembly\_fragment"  
misc\_feature 35757..40409  
/note="assembly\_fragment"  
misc\_feature 40510..46096  
/note="assembly\_fragment"  
misc\_feature 46197..50343  
/note="assembly\_fragment"  
misc\_feature 50444..54901  
/note="assembly\_fragment"  
misc\_feature 55002..59436  
/note="assembly\_fragment"  
misc\_feature 59537..66852  
/note="assembly\_fragment"  
misc\_feature 66953..75809  
/note="assembly\_fragment"  
misc\_feature 75910..85009  
/note="assembly\_fragment"  
misc\_feature 85110..94166  
/note="assembly\_fragment"  
misc\_feature 94267..104445

Query Match 34.7% Score 104; DB 2: Length 202689;  
Best Local Similarity 72.8% Pred. No. 4.3e-16;  
Matches 134; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 1 GGGGGCCGTATCCAGCTACTCAGAGGCTGAGGAGGATCCGCGAGGCTTGCA 60  
DB 14920 GGGGGCCGTATCCAGCTACTCAGAGGCTGAGGAGGATCCGCGAGGCTTGCA 14979  
QY 61 GATCTGCTGAGCTGGAGGTTGAGGCTACAGTAAGCAAGATCATGCTGATTAATCTTC 120

DB 14980 GAATCGCTTGAACCCAGAGCGAGGTTGCAATGACCGAGATCGGACCTGCCTCC 15039  
QY 121 AGCCGTGGCGCAGAAAGTGAGACCGTAAACAAAAAATTTAAAGAAATTTAG 180  
DB 15040 AGCCGTGGCGCAGAAAGTGAGACCTGATCTCAAAATTAACGATTAAAGTAAATAAA 15099  
QY 181 ATCA 184  
DB 15100 ATTA 15103

RESULT 13  
AC146464  
LOCUS  
DEFINITION Saimiri sciureus clone CH254-84A11, WORKING DRAFT SEQUENCE.  
ACCESSION AC146464  
VERSION AC146464.1 GI:33636782  
KEYWORDS HTG; HTGS PHASE2; HTGS DRAFT.  
SOURCE Saimiri sciureus (common squirrel monkey)  
ORGANISM Saimiri sciureus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Cebinae;  
Saimiri.

REFERENCE 1 (bases 1 to 202533)  
AUTHORS Cheng, J.-F., Hamilton, M., Peng, Y., Mukherjee, S., Hosseini, R.,  
Peng, Z., Malinov, I. and Rubin, E.M.  
DIRECT SUBMISSION  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 202533)  
AUTHORS Cheng, J.-F., Hamilton, M., Peng, Y., Mukherjee, S., Hosseini, R.,  
Peng, Z., Malinov, I. and Rubin, E.M.  
DIRECT SUBMISSION  
JOURNAL Submitted (14-AUG-2003) Genome Sciences, Lawrence Berkeley National  
Laboratory, 1 Cyclotron Rd., Berkeley, CA 94720, USA  
REFERENCE 3 (bases 1 to 202533)  
AUTHORS Cheng, J.-F., Hamilton, M., Peng, Y., Mukherjee, S., Hosseini, R.,  
Peng, Z., Malinov, I. and Rubin, E.M.  
DIRECT SUBMISSION  
JOURNAL Submitted (19-AUG-2003) Genome Sciences, Lawrence Berkeley National  
Laboratory, 1 Cyclotron Rd., Berkeley, CA 94720, USA  
COMMENT  
Sequence Produced by Berkeley PGA  
Web site: <http://pga.lbl.gov>  
Center Code: PGASERK  
Center Project Name: S030  
Bac Clone Name: CH254-84A11

This sequence has been compared to sequences of other species  
using VISTA (<http://www-gsd.lbl.gov/VISTA/>). The results can be  
viewed at:  
[http://pga.lbl.gov/cgi-bin/search\\_cvcgdt?type=ABCG5](http://pga.lbl.gov/cgi-bin/search_cvcgdt?type=ABCG5)

The order-orientation of the draft sequence was accomplished by  
using:  
Avid (<http://baboon.math.berkeley.edu/avid/>),  
Lagan (<http://lagan.stanford.edu/>) and paired end information.

Funding agent: Programs for Genomic Applications (NHLBI)

#### Summary Statistics:

Sequencing vector: Plasmid, pUC18  
Chemistry: Dye-terminator Big Dye  
Assembly Program: Phrap version 0.990329.

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 1 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.



## FEATURES

1 202533: contig of 202533 bp in length.  
Location/Qualifiers  
1. 202533  
/organism="Saimiri sciureus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9521"  
/clone="CH254-84A11"

## ORIGIN

Query Match 34.5%; Score 103.6; DB 2; Length 202533;  
Best Local Similarity 68.9%; Pred. No. 5.4e-16;  
Matches 142; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 5 GCGTGAATCCAGCTACTGAGAGGCTGAGGCGAGAGATCCGCGAGCCTGGCAGATC 64  
DB 87655 GCGTGAATCCAGCTACTGAGAGGCTGAGGCGAGAGATCCGCGAGCCTGGCAGATC 87714  
QY 65 TGCGTGAAGCTGGAGGTTGAGGCTGACGTAAGCCAGATCATGCCAGTACTTCAGCC 124  
DB 87715 CGCTTGAACCTGGAGGATCGAGGCTGACGTAAGCTGATCATGCCAGTACTTCAGCC 87774  
QY 125 TGGGCGAAGAGCTGAGAGCTGACCAAAAAAATTTAAAAAAGAAATTTAGATCA 164  
DB 87775 TGGGCGAAGAGAGCTGAGAGCTGACCAAAAAAATTTAAAAAAGAAATTTAGATCA 87834  
QY 185 AGATCCAACTGTAAAAAGTGGCTTA 210  
DB 87835 TCCTTATCTTAAATCAAGTTCAA 87860

## RESULT 14

AC146787/c AC146787 178016 bp DNA linear HTG 03-OCT-2003  
LOCUS Aotus nancymae clone CH258-323A5, WORKING DRAFT SEQUENCE, 4  
DEFINITION  
ORDERED PIECES.  
AC146787  
AC146787.1 GI:37497135  
VERSION HTG; HTGS PHASE2; HTGS DRAFT.  
KEYWORDS Aotus nancymae (Ma's Night monkey)  
SOURCE Aotus nancymae  
ORGANISM Aotus nancymae  
REFERENCE 1 (bases 1 to 178016)  
AUTHORS Cheng, J.-F., Hamilton, M., Peng, Y., Mukherjee, S., Hosseini, R., Peng, Z., Malinov, I. and Rubin, E.M.  
TITLE Unpublished  
JOURNAL Direct Submission  
AUTHORS 2 (bases 1 to 178016)  
TITLE Cheng, J.-F., Hamilton, M., Peng, Y., Mukherjee, S., Hosseini, R., Peng, Z., Malinov, I. and Rubin, E.M.  
JOURNAL Direct Submission  
AUTHORS Submitted (03-OCT-2003) Genome Sciences, Lawrence Berkeley National Laboratory, 1 Cyclotron Rd., Berkeley, CA 94720, USA  
COMMENT  
Sequence Produced by Berkeley PGK  
Web site: <http://pga.lbl.gov>  
Center Code: PGABERK  
Center Project Name: W010  
Bac Clone Name: CH258-323A5

This sequence has been compared to sequences of other species using Vindex (<http://www.gsc.lbl.gov/VINDEX>). The results can be viewed at:  
[http://pga.lbl.gov/cgi-bin/search\\_cvgcd?type=ncvalue=ABCG5](http://pga.lbl.gov/cgi-bin/search_cvgcd?type=ncvalue=ABCG5)

The order-orientation of the draft sequence was accomplished by using:  
Avid (<http://baboon.mach.berkeley.edu/avid/>),  
Lagan (<http://lagan.stanford.edu/>) and paired end information.

Funding agent: Programs for Genomic Applications (NLBBI)

Summary Statistics:

Sequencing vector: plasmid, pUC18  
Chemistry: Dye-terminator Big Dye  
Assembly program: Phrap version 0.990329.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 4 contigs. Gaps between the contigs  
\* are represented as runs of 'N'. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 32150: contig of 32150 bp in length  
\* 32151 32250: gap of unknown length  
\* 32251 56222: contig of 23972 bp in length  
\* 56223 56322: gap of unknown length  
\* 56323 173105: contig of 116783 bp in length  
\* 173106 173205: gap of unknown length  
\* 173206 178016: contig of 4811 bp in length.  
Location/Qualifiers  
1. 178016  
/organism="Aotus nancymae"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:37293"  
/clone="CH258-323A5"

## ORIGIN

Query Match 34.4%; Score 103.2; DB 2; Length 178016;  
Best Local Similarity 69.1%; Pred. No. 6.8e-16;  
Matches 141; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 7 CTGTATCCAGCTCTCAGAGGCTGAGGCGAGATCCGCGAGCCTGGCAGATCTG 66  
DB 40241 CTGTATCCAGCTCTCAGAGGCTGAGGCGAGATCCGCGAGCCTGGCAGATCTG 40182  
QY 67 CCTGAGCCTGGAGGTTGAGGCTGACGTAAGCAAGATCATGCCAGTACTTCAAGCTG 126  
DB 40181 CCTGAGCCTGGAGGTTGAGGCTGACGTAAGCAAGATCATGCCAGTACTTCAAGCTG 40122  
QY 127 GCGCAAGAGTGAGACCTGTAACAAAAAATTTAAAAAAGAAATTTAGATCAAG 186  
DB 40121 GCGCAAGAGTGAGACCTGTAACAAAAAATTTAAAAAAGAAATTTAGATCAAG 40062  
QY 187 ATCCAACTGTAAAAAGTGGCTTA 210  
DB 40061 ATCCAACTGTAAAAAGTGGCTTA 40038

## RESULT 15

AC007601/c AC007601 161864 bp DNA linear PRI 14-MAR-2003  
LOCUS Homo sapiens chromosome 16 clone RP11-276H1, complete sequence.  
DEFINITION  
AC007601  
AC007601.7 GI:28372590  
VERSION  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 161864)  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
TITLE DOE Joint Genome Institute, Stanford Human Genome Center and Los  
JOURNAL Alamos National Laboratory.  
AUTHORS Direct Submission  
Unpublished

2 (bases 1 to 161864)  
TITLE Bruce, D., Mundt, M., Doggett, N., Munk, C., Saunders, E., Robinson, D.,  
JONES, M., Buckingham, J., Chasteen, L., Thompson, S., Goodwin, L.,  
Bryant, J., Teemer, J., Weinke, L., Longmire, J., White, S., Tatam, O.,  
Campbell, C., Fawcett, J., Maltbie, M., Bussod, M., Sutherland, R.,  
McMurry, K., Han, C. and Deaven, L.  
JOURNAL Direct Submission  
Submitted (20-MAR-1999) Center for Human Genome Studies, DOE Joint  
Genome Institute, Los Alamos National Laboratory, MS M888, Los



REFERENCE Alamos, NM 87545, USA  
AUTHORS 3 (bases 1 to 161864)  
TITLE DOE Joint Genome Institute.  
JOURNAL Direct Submission  
SUBMITTED (29-MAR-2002) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
AUTHORS 4 (bases 1 to 161864)  
TITLE DOE Joint Genome Institute.  
JOURNAL Direct Submission  
SUBMITTED (13-FEB-2003) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
AUTHORS 5 (bases 1 to 161864)  
TITLE DOE Joint Genome Institute, Stanford Human Genome Center and Los  
Alamos National Laboratory.  
JOURNAL Direct Submission  
SUBMITTED (14-MAR-2003) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA  
COMMENT On Feb 13, 2003 this sequence version replaced gi:19807785.  
Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center and Los Alamos  
National Laboratory  
www.sngc.stanford.edu  
Quality: Phrap Quality >=40 100% of Sequence;  
Estimated Total Number of Errors is 0.  
Location/Qualifiers  
1. 161864  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="RP11-276H1"

## FEATURES

## SOURCE

## ORIGIN

Query Match 34.1%; Score 102.4; DB 9; Length 161864;  
Best Local Similarity 72.3%; Pred.No. 1.1e-15;  
Matches 133; Conservative 0; Mismatches 51; Indels 0; Gaps 0;  
QY 1 GGGGGCTGTATCCCACTACTCAGAGGCTGAGGAGGAGATCCCGGAGCTGGCA 60  
DB 87397 GGGTGCCTGTATCCCACTACTCAGAGGCTGAGGAGGAGATCCCGGAGCTGGCA 87338  
QY 61 GATCGCTGAGCGCTGAGGCTGAGGCTAGAGCAAGATCATGCCAGTACTTC 120  
DB 87337 GATGCTTGAACCCAGAGGCGGAGGTTGCAATGAGCCGATGCTGCACTGCC 87278  
QY 121 AGCTGGGCGACAAAGTAGACCGTAAACAAAAAATTTAAAAAAGAAATTTAG 180  
DB 87277 AGCTGGGCGACAAAGTAGAGTATCTCAAAATTAACGATTAACTAAGTAATAAA 87218  
QY 181 ATCA 184  
DB 87217 ATTA 87214

Search completed: March 12, 2004, 21:22:23  
Job time : 948.481 secs







CC The present invention describes an activation-induced cytidine deaminase  
CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and has  
CC cytidine activity similar to APOBEC-1. AID has anti-allergic, anti-inflammatory,  
CC anti-inflammatory, ophthalmological, anti-HIV and dermatological activities,  
CC and can be used in gene therapy. AID polynucleotides are useful in  
CC methods for identifying drugs for the treatment of B cell associated  
CC immune system disorders, immunodeficiency diseases and allergies, such as  
CC immunoglobulin A (IgA) deficiency disease, IgA nephritis, gamma-  
CC globulinemia, atopic dermatitis, allergic colitis, asthma, food allergy,  
CC drug allergy, allergic rhinitis, Rosen disease, DiGeorge disease, ataxia  
CC telangiectasia, common variable immunodeficiency disorder, MHC (major  
CC histocompatibility class II deficiency disease, AIDS (auto  
CC immunodeficiency syndrome), elevated IgE deficiency disease, and IgG subclass  
CC selection disorder. The DNA sequences encoding AID may be used for gene  
CC therapy and the antibodies to the AID protein may be used for diagnosis  
CC and treatment of these disorders. The present sequence represents a  
CC genomic DNA sequence of human AID  
XX

SO Sequence 6564 BP; 1909 A; 1358 C; 1383 G; 1914 T; 0 U; 0 Other;

Query Match 100.0%; Score 300; DB 3; Length 6564;

Best Local Similarity 100.0%; Pred. No. 2.3e-63;

Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GGGGGCCTGTAATCCAGCTACTCAGAGAGCTGAGGCGAGAGATCCCGGAGCTGGCA 60

1 GGGGGCCTGTAATCCAGCTACTCAGAGAGCTGAGGCGAGAGATCCCGGAGCTGGCA 60

61 GATCGGCTGAGGCTGGAGGTTGAGGCTGACAGTAAAGATGATGAGTATATCTTC 120

61 GATCGGCTGAGGCTGGAGGTTGAGGCTGACAGTAAAGATGATGAGTATATCTTC 120

121 AGCGTGGCGCAAGTGAACCGTAAACAAAAAATTTAAAAAAGAAATTTAG 180

121 AGCGTGGCGCAAGTGAACCGTAAACAAAAAATTTAAAAAAGAAATTTAG 180

181 ATCAAGATCCAACTGTAATAAGTGGCTTAAACCAATTAAGATTTGAGTTATTC 240

181 ATCAAGATCCAACTGTAATAAGTGGCTTAAACCAATTAAGATTTGAGTTATTC 240

241 TGAGGCGAAGAGAACCATCGGGGCTTTCAGCATGGGAATGGATGTCACCTGGT 300

241 TGAGGCGAAGAGAACCATCGGGGCTTTCAGCATGGGAATGGATGTCACCTGGT 300

241 TGAGGCGAAGAGAACCATCGGGGCTTTCAGCATGGGAATGGATGTCACCTGGT 300

241 TGAGGCGAAGAGAACCATCGGGGCTTTCAGCATGGGAATGGATGTCACCTGGT 300

241 TGAGGCGAAGAGAACCATCGGGGCTTTCAGCATGGGAATGGATGTCACCTGGT 300

241 TGAGGCGAAGAGAACCATCGGGGCTTTCAGCATGGGAATGGATGTCACCTGGT 300

241 TGAGGCGAAGAGAACCATCGGGGCTTTCAGCATGGGAATGGATGTCACCTGGT 300

241 TGAGGCGAAGAGAACCATCGGGGCTTTCAGCATGGGAATGGATGTCACCTGGT 300

241 TGAGGCGAAGAGAACCATCGGGGCTTTCAGCATGGGAATGGATGTCACCTGGT 300

241 TGAGGCGAAGAGAACCATCGGGGCTTTCAGCATGGGAATGGATGTCACCTGGT 300

241 TGAGGCGAAGAGAACCATCGGGGCTTTCAGCATGGGAATGGATGTCACCTGGT 300

241 TGAGGCGAAGAGAACCATCGGGGCTTTCAGCATGGGAATGGATGTCACCTGGT 300

241 TGAGGCGAAGAGAACCATCGGGGCTTTCAGCATGGGAATGGATGTCACCTGGT 300

241 TGAGGCGAAGAGAACCATCGGGGCTTTCAGCATGGGAATGGATGTCACCTGGT 300

241 TGAGGCGAAGAGAACCATCGGGGCTTTCAGCATGGGAATGGATGTCACCTGGT 300

241 TGAGGCGAAGAGAACCATCGGGGCTTTCAGCATGGGAATGGATGTCACCTGGT 300

28-MAR-2000; 2000MO-JP001918.

XX 29-MAR-1999; 99GP-00087192.  
PR 24-JUN-1999; 99GP-00178999.  
PR 27-DEC-1999; 99GP-00371382.  
XX (NIBS) JAPAN TOBACCO INC.  
PA (HONU/) HONJO T.  
XX

PI Honjo T, Muramatsu M,  
XX WPI; 2000-611715/58.  
DR

XX Nucleic acid encoding activation induced cytidine deaminase, useful as a  
PT target for drug development for immune-related diseases including  
PT allergies.  
XX

PS Claim 17; Page 163-170; 174pp; Japanese.

XX The present invention describes an activation-induced cytidine deaminase  
CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and has  
CC cytidine activity similar to APOBEC-1. AID has anti-allergic, anti-inflammatory,  
CC anti-inflammatory, ophthalmological, anti-HIV and dermatological activities,  
CC and can be used in gene therapy. AID polynucleotides are useful in  
CC methods for identifying drugs for the treatment of B cell associated  
CC immune system disorders, immunodeficiency diseases and allergies, such as  
CC immunoglobulin A (IgA) deficiency disease, IgA nephritis, gamma-  
CC globulinemia, atopic dermatitis, allergic colitis, asthma, food allergy,  
CC drug allergy, allergic rhinitis, Rosen disease, DiGeorge disease, ataxia  
CC telangiectasia, common variable immunodeficiency disorder, MHC (major  
CC histocompatibility class II deficiency disease, AIDS (auto  
CC immunodeficiency syndrome), elevated IgE deficiency disease, and IgG subclass  
CC selection disorder. The DNA sequences encoding AID may be used for gene  
CC therapy and the antibodies to the AID protein may be used for diagnosis  
CC and treatment of these disorders. The present sequence represents a  
CC genomic DNA sequence of human AID  
XX

SO Sequence 11204 BP; 3305 A; 2273 C; 2373 G; 3253 T; 0 U; 0 Other;

Query Match 99.5%; Score 298.4; DB 3; Length 11204;

Best Local Similarity 99.7%; Pred. No. 6.5e-63;

Matches 299; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 GGGGGCCTGTAATCCAGCTACTCAGAGAGCTGAGGCGAGAGATCCCGGAGCTGGCA 60

5216 GGGGGCCTGTAATCCAGCTACTCAGAGAGCTGAGGCGAGAGATCCCGGAGCTGGCA 5275

61 GATCGGCTGAGGCTGGAGGTTGAGGCTGACAGTAAAGATGATGAGTATATCTTC 120

5276 GATCGGCTGAGGCTGGAGGTTGAGGCTGACAGTAAAGATGATGAGTATATCTTC 5335

121 AGCGTGGCGCAAGTGAACCGTAAACAAAAAATTTAAAAAAGAAATTTAG 180

5336 AGCGTGGCGCAAGTGAACCGTAAACAAAAAATTTAAAAAAGAAATTTAG 5395

181 ATCAAGATCCAACTGTAATAAGTGGCTTAAACCAATTAAGATTTGAGTTATTC 240

5396 ATCAAGATCCAACTGTAATAAGTGGCTTAAACCAATTAAGATTTGAGTTATTC 5455

241 TGAGGCGAAGAGAACCATCGGGGCTTTCAGCATGGGAATGGATGTCACCTGGT 300

5456 TGAGGCGAAGAGAACCATCGGGGCTTTCAGCATGGGAATGGATGTCACCTGGT 5515

5456 TGAGGCGAAGAGAACCATCGGGGCTTTCAGCATGGGAATGGATGTCACCTGGT 5515

5456 TGAGGCGAAGAGAACCATCGGGGCTTTCAGCATGGGAATGGATGTCACCTGGT 5515

5456 TGAGGCGAAGAGAACCATCGGGGCTTTCAGCATGGGAATGGATGTCACCTGGT 5515

5456 TGAGGCGAAGAGAACCATCGGGGCTTTCAGCATGGGAATGGATGTCACCTGGT 5515

5456 TGAGGCGAAGAGAACCATCGGGGCTTTCAGCATGGGAATGGATGTCACCTGGT 5515

5456 TGAGGCGAAGAGAACCATCGGGGCTTTCAGCATGGGAATGGATGTCACCTGGT 5515

5456 TGAGGCGAAGAGAACCATCGGGGCTTTCAGCATGGGAATGGATGTCACCTGGT 5515

5456 TGAGGCGAAGAGAACCATCGGGGCTTTCAGCATGGGAATGGATGTCACCTGGT 5515

5456 TGAGGCGAAGAGAACCATCGGGGCTTTCAGCATGGGAATGGATGTCACCTGGT 5515

5456 TGAGGCGAAGAGAACCATCGGGGCTTTCAGCATGGGAATGGATGTCACCTGGT 5515

5456 TGAGGCGAAGAGAACCATCGGGGCTTTCAGCATGGGAATGGATGTCACCTGGT 5515

5456 TGAGGCGAAGAGAACCATCGGGGCTTTCAGCATGGGAATGGATGTCACCTGGT 5515

5456 TGAGGCGAAGAGAACCATCGGGGCTTTCAGCATGGGAATGGATGTCACCTGGT 5515

5456 TGAGGCGAAGAGAACCATCGGGGCTTTCAGCATGGGAATGGATGTCACCTGGT 5515

5456 TGAGGCGAAGAGAACCATCGGGGCTTTCAGCATGGGAATGGATGTCACCTGGT 5515

04-DEC-2002 (first entry)  
DNA encoding human translocation del(12p) protein #1.



KM Chromosome aberration; oncogenic fusion protein; cancer;  
KM Proliferative disease; cellular protein isoform; heat shock protein 90;  
KM HSP-90; rheumatoid arthritis; cancer; haematopoietic disorder;  
KM T cell lymphoma; B cell lymphoma; chronic myeloid leukaemia; CML;  
KM acute myeloid leukaemia; AML; chronic myelomonocytic leukaemia; CMML;  
KM papillary thyroid carcinoma; Ewing's sarcoma; melanoma; liposarcoma;  
KM rhabdomyosarcoma; synovial sarcoma; viral infection; gene; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200269900-A2.  
XX  
PD 12-SEP-2002.  
XX  
PF 01-MAR-2002; 2002WO-US006518.  
XX  
PR 01-MAR-2001; 2001US-0272751P.  
XX  
PA (CONF-) CONFORMA THERAPEUTICS CORP.  
XX  
PI Fritz LC, Burrows FJ;  
XX  
XX WPI; 2002-698710/75.  
DR P-PSDB; ABG95082.  
XX  
XX Treating genetically-defined disease associated with chromosomal  
PT aberrations yielding oncogenic fusion proteins, e.g. cell proliferative  
PT diseases, involves administering an inhibitor of heat shock protein 90.  
XX  
PS Disclosure; Page 242-245; 389pp; English.  
XX  
XX The invention describes a method of treating genetically-defined disease  
CC associated with chromosomal aberrations yielding oncogenic fusion  
CC proteins (I), treating cancerous cells containing (I) in a heterogeneous  
CC cell population, treating proliferative diseases associated with mutant  
CC protein or cellular protein isoform (II) dependent on heat shock protein  
CC (HSP)-90, or selectively treating cells expressing (II) involving  
CC administering HSP90-inhibitor. The method is useful for treating  
CC genetically-defined disease with chromosomal aberration yielding  
CC oncogenic fusion protein, treating cancerous cells containing fusion  
CC protein in heterogeneous cell population, treating proliferative disease  
CC (e.g. rheumatoid arthritis or cancer) associated with mutant protein or  
CC cellular protein isoform dependent on heat shock protein (HSP)-90 (e.g.  
CC p53), or selectively treating cells expressing mutant protein or cellular  
CC protein isoform in a patient heterozygous for (II). The method is useful  
CC for treating a disease e.g. haematopoietic disorder such as T or B cell  
CC lymphoma, chronic myeloid leukaemia (CML), APL, ALL, NHL and CMML,  
CC or a disease characterised by a solid tumour such as papillary thyroid  
CC carcinoma, Ewing's sarcoma, melanoma, liposarcoma, rhabdomyosarcoma and  
CC synovial sarcoma. The method is also useful for treating viral  
CC infections. This represents the DNA sequence of a chromosome aberration  
XX  
SQ Sequence 11204 BP; 3305 A; 2273 C; 2373 G; 3253 T; 0 U; 0 Other;  
Query Match 99.5%; Score 298.4; DB 6; Length 11204;  
Best Local Similarity 99.7%; Pred. No. 6; se-63;  
Matches 299; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 GGGGGCTGTATATCCAGCTACTCAGAGGCTGAGGAGGATCCGGAGCTGGCA 60  
DB 5216 GGGGCTGTATATCCAGCTACTCAGAGGCTGAGGAGGATCCGGAGCTGGCA 5275  
QY 61 GATTCGCTGAGCTGGAGGCTGAGGCTACAGTAAGCCAGATCATGCGAGTACTTC 120  
DB 5276 GATTCGCTGAGCTGGAGGCTGAGGCTACAGTAAGCCAGATCATGCGAGTACTTC 5335  
QY 121 AGCTGGGCGCAAGAGGAGGAGGCTGAGGCTACAGTAAGCCAGATCATGCGAGTACTTC 180  
DB 5336 AGCTGGGCGCAAGAGGAGGAGGCTGAGGCTACAGTAAGCCAGATCATGCGAGTACTTC 5395  
QY 181 ATCAAGATCCACTGTAAAAAGTGCGCTAACAACCATTTAAAGAGTTGAGTTATTC 240

DB 5396 ATCAAGATCCACTGTAAAAAGTGCGCTAACAACCATTTAAAGAGTTGAGTTATTC 5455  
QY 241 TGCGAGGAGAGAGACATCAGGAGGCTTCAGCATGGGATGGCATGCACTGTGT 300  
DB 5456 TGCGAGGAGAGAGACATCAGGAGGCTTCAGCATGGGATGGCATGCACTGTGT 5515  
RESULT 4  
AAL50814/C  
ID AAL50814 standard; DNA; 79528 BP.  
XX  
XX AAL50814;  
XX  
XX 30-JAN-2003 (first entry)  
XX  
DE Human cancer status prediction method-related DNA sequence #6.  
XX  
XX Human; gene therapy; cancer status prediction; cancer; ds;  
XX  
XX cancer malignancy evaluation; drug design; antisense nucleic acid.  
XX  
OS Homo sapiens.  
XX  
XX WO200272828-A1.  
XX  
XX 19-SEP-2002.  
XX  
XX 07-MAR-2002; 2002WO-JP002153.  
XX  
XX 14-MAR-2001; 2001JP-00073063.  
XX  
XX 06-APR-2001; 2001JP-00108503.  
XX  
XX 02-AUG-2001; 2001JP-00234807.  
XX  
XX (DNAC-) DNA CHIP RES INC.  
XX  
XX (HISF) HITACHI SOFTWARE ENG CO LTD.  
XX  
XX Kato K, Iwao K, Noguchi S, Maroba R;  
XX  
XX WPI; 2002-713517/77.  
XX  
XX Computer-aided statistical method for predicting cancer, applicable in  
PT gene therapy for evaluating cancer malignancy with data for use in drug  
PT design.  
XX  
PS Disclosure; Page 84-131; 182pp; Japanese.  
XX  
XX The invention comprises a method for predicting cancer status. The method  
XX involves: measuring expression doses of genes obtained from specimens;  
XX selecting at least one gene as the gene for an assay; using the  
XX measurement results on expression doses of the selected genes for  
XX multivariate analysis; and classifying the specimens in analogous groups  
XX CC with results of the multivariate analysis on expression patterns of the  
XX CC genes. The method of the invention is useful for predicting cancer, which  
XX CC is applicable in gene therapy for evaluating cancer malignancy with data  
XX CC for use in drug design (e.g. antisense nucleic acids for use in gene  
XX CC therapy to treat cancer). The present DNA sequence represents a human  
XX CC nucleic acid of the invention  
XX  
SQ Sequence 79528 BP; 19015 A; 20270 C; 20468 G; 19775 T; 0 U; 0 Other;  
Query Match 36.4%; Score 109.2; DB 6; Length 79528;  
Best Local Similarity 72.7%; Pred. No. 1; se-16;  
Matches 141; Conservative 0; Mismatches 53; Indels 0; Gaps 0;  
QY 1 GGGGGCTGTATATCCAGCTACTCAGAGGCTGAGGAGGATCCGGAGCTGGCA 60  
DB 64094 GGGGCTGTATATCCAGCTACTCAGAGGCTGAGGAGGATCCGGAGCTGGCA 64035  
QY 61 GATTCGCTGAGCTGGAGGCTGAGGCTACAGTAAGCCAGATCATGCGAGTACTTC 120  
DB 64094 GATTCGCTGAGCTGGAGGCTGAGGCTACAGTAAGCCAGATCATGCGAGTACTTC 63975  
QY 121 AGCTGGGCGCAAGAGGAGGAGGCTGAGGCTACAGTAAGCCAGATCATGCGAGTACTTC 180



Db 63974 AGCCTGGAGACGAAACGAGACTCCATTCCTCAAAAAAAAAATAAATAATAATAAAAAATA 63911

Oy 181 ATCAAGTCCA ACT 194  
| | | | |  
Db 63914 ATAATTAAATTAAT 63901

RESULT 5  
ABQ88117  
ID ABQ88117 standard; cDNA; 129722 BP.  
XX ABQ88117;  
XX AC  
XX XX  
XT 18-SEP-2002 (first entry)  
DE Human osteoblast differentiation related cDNA SEQ ID NO 24.

XX Human; osteoblast; stem cell differentiation; bone tissue deposition;  
KW osteoporosis; osteopathia; ss.  
XX Homo sapiens.  
CS WO200250301-A2.  
XX PD 27-JUN-2002.  
XX PP 18-DEC-2001; 2001WO-US048276.  
XX PR 18-DEC-2000; 2000US-0255882P.  
XX PR 24-APR-2001; 2001US-0285691P.  
PA (GENE-) GENE LOGIC INC.  
PP (PROC) PROCTER & GAMBLE CO.

XX Ji D, Axelrod DW, Cook JS, Jaiswal N, Einstein R, Houghton A;  
PI Mertz L;  
DR WPI; 2002-557663/59.

XX Use of genes and their expression profiles associated with osteoblast  
PT differentiation for screening modulators bone formation, for diagnosing  
PT or treating e.g. osteoporosis, or as markers for the differentiation  
PT process.

PS Claim 1; SEQ ID NO 24; 78bp + Sequence Listing; English.

XX The invention relates to genes and their expression profiles are used  
CC for: (a) screening modulators of precursor stem cell differentiation into  
CC osteoblasts, or bone tissue deposition; (b) diagnosing abnormal  
CC deposition of bone tissue, abnormal rate of osteoblast formation OR  
CC osteoporosis; or (c) treating or monitoring treatment of the conditions  
CC cited in (b); or monitoring the progression of bone tissue deposition.  
CC Specific conditions include postmenopausal osteoporosis, glucocorticoid  
CC osteoporosis or male osteoporosis, osteopenia, osteodystrophy, drug-  
CC induced abnormalities in bone formation or bone loss, conditions that  
CC involve altered bone metabolism (e.g. idiopathic juvenile osteoporosis),  
CC skeletal disease linked to breast cancer, mastocytosis, Fanconi syndrome  
CC or fibrous dysplasia. The present sequence is that of an osteoblast  
CC differentiation associated cDNA marker of the invention. Note: The  
CC sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
at ftp.wipo.int/pub/published\_pct\_sequences

SQ Sequence 129722 BP; 33298 A; 32697 C; 31415 G; 32312 T; 0 U; 0 Other;

Query Match 32.7%; Score 98; DB 6; Length 129722;  
Best Local Similarity 76.4%; Pred. No. 8.le-14;  
Matches 133; Conservative 0; Mismatches 40; Indels 1; Gaps 1;

Oy 1 GGGGGCGCTGTATCCAGCTACTGAGGAGGCTGAGGAGGAGTGATCCGCAGCGCTTGACA 60  
| | | | |  
Db 49787 GGCGCGCTGTACTGCCAGCTACTGGAGAGCTGAGGAGCATGTACTGTGAAGCAGG-A 49844

Oy	61	GATTCGCTGAGCTGGAGGATTGAGGCTACAGTAAAGCCAGATCATGCGAGTACTTC	120
Dd	49946	GAATGGCTTAACCTGGAGAGCAGATTTCGATGAGCCGAGATGTGTCCACTGCACCTCC	49905
Oy	121	AGCCTGGCGCACAAGTGAGACCGGTACCAAAAAAAAAAATTTAAATAAAAGAA	174
Dd	49906	AGCCTGGGTACAGAGTGAACCTTGTCTCAAAAAAAAAAAAAAAAAAAAAAGAA	49959
RESULT 6			
ID	AAI03615		
XX	AAI03615	standard; DNA; 32134 BP.	
AC	AAI03615;		
XX			
DT	21-NOV-2001	(first entry)	
XX			
DE	Human reproductive system related antigen DNA SEQ ID NO: 6303.		
XX	Human; reproductive system related antigen; reproductive system disorder;		
KM	Human; gene therapy; ds.		
XX			
OS	Homo sapiens.		
PN	MO200155320-A2.		
XX			
PD	02-AUG-2001.		
PF	17-JAN-2001; 2001WO-US001339.		
XX			
PR	31-JAN-2000; 2000US-0179065P.		
PR	04-FEB-2000; 2000US-0180628P.		
PR	24-FEB-2000; 2000US-0184664P.		
PR	02-MAR-2000; 2000US-0186350P.		
PR	16-MAR-2000; 2000US-0189874P.		
PR	17-MAR-2000; 2000US-0190076P.		
PR	18-APR-2000; 2000US-0198123P.		
PR	19-MAY-2000; 2000US-0205515P.		
PR	07-JUN-2000; 2000US-0209467P.		
PR	28-JUN-2000; 2000US-0214886P.		
PR	30-JUN-2000; 2000US-0215135P.		
PR	07-JUL-2000; 2000US-0216647P.		
PR	07-JUL-2000; 2000US-0216880P.		
PR	11-JUL-2000; 2000US-0217487P.		
PR	11-JUL-2000; 2000US-0217496P.		
PR	14-JUL-2000; 2000US-0218290P.		
PR	26-JUL-2000; 2000US-0220963P.		
PR	26-JUL-2000; 2000US-0220964P.		
PR	14-AUG-2000; 2000US-0224518P.		
PR	14-AUG-2000; 2000US-0224519P.		
PR	14-AUG-2000; 2000US-0225213P.		
PR	14-AUG-2000; 2000US-0225214P.		
PR	14-AUG-2000; 2000US-0225266P.		
PR	14-AUG-2000; 2000US-0225267P.		
PR	14-AUG-2000; 2000US-0225268P.		
PR	14-AUG-2000; 2000US-0225270P.		
PR	14-AUG-2000; 2000US-0225477P.		
PR	14-AUG-2000; 2000US-0225757P.		
PR	14-AUG-2000; 2000US-0225758P.		
PR	14-AUG-2000; 2000US-0225759P.		
PR	18-AUG-2000; 2000US-0226279P.		
PR	22-AUG-2000; 2000US-0226681P.		
PR	22-AUG-2000; 2000US-0226686P.		
PR	22-AUG-2000; 2000US-0227183P.		
PR	23-AUG-2000; 2000US-0227093P.		
PR	30-AUG-2000; 2000US-0228924P.		
PR	01-SEP-2000; 2000US-0229287P.		
PR	01-SEP-2000; 2000US-0229343P.		
PR	01-SEP-2000; 2000US-0229344P.		
PR	01-SEP-2000; 2000US-0229345P.		
PR	05-SEP-2000; 2000US-0229509P.		
PR	05-SEP-2000; 2000US-0229513P.		
PR	06-SEP-2000; 2000US-0230437P.		



[illegible]

PR	17-NOV-2000;	2000US-0249244P.
PR	17-NOV-2000;	2000US-0249245P.
PR	17-NOV-2000;	2000US-0249264P.
PR	17-NOV-2000;	2000US-0249265P.
PR	17-NOV-2000;	2000US-0249297P.
PR	17-NOV-2000;	2000US-0249299P.
PR	17-NOV-2000;	2000US-0249300P.
PR	01-DEC-2000;	2000US-0250180P.
PR	01-DEC-2000;	2000US-0250391P.
PR	05-DEC-2000;	2000US-0251030P.
PR	05-DEC-2000;	2000US-0251988P.
PR	05-DEC-2000;	2000US-0256719P.
PR	06-DEC-2000;	2000US-0251479P.
PR	08-DEC-2000;	2000US-0251856P.
PR	08-DEC-2000;	2000US-0251868P.
PR	08-DEC-2000;	2000US-0251869P.
PR	08-DEC-2000;	2000US-0251989P.
PR	08-DEC-2000;	2000US-0251990P.
PR	11-DEC-2000;	2000US-0254057P.
PR	05-JAN-2001;	2001US-0259678P.
XX	(HUMA-) HUMAN GENOME SCI INC.	
PA		
PI	Rosen CA, Barash SC, Ruben SM;	
XX		
DR	WPI; 2001-465570/50.	
XX		
PT	Isolated nucleic acid molecule encoding a reproductive system antigen is used in preventing, treating or ameliorating a medical condition.	
XX		
PS	Disclosure; SEQ ID NO 6303; 1297bp + Sequence Listing; English.	
XX		
CC	The present invention provides the protein and coding sequences of a number of human reproductive system related antigens. These can be used in the prevention and treatment of reproductive system disorders, including cancer. The present sequence is a genomic sequence encoding a protein of the invention	
CC		
CC		
CC		
SQ	Sequence 32134 BP; 9171 A; 6268 C; 6931 G; 9764 T; 0 U; 0 Other;	
XX		
Query Match	31.3%; Score 94; DB 4; Length 32134;	
Best Local Similarity	71.3%; Pred. No. 5.5e-13;	
Matches 124; Conservative	0; Mismatches 50; Indels 0; Gaps 0;	
OY	1 GGGGGCGCTGAATCCAGCTACTCAGGAGGCTGAGGCAGAGGANTCCGGGACCTTGCA 60	
DB	12455 GGGGACCTGTACTCCAGCTACTTCGGGAGGCTGAGGCAGAGGAAATGCATGAGGAGGA 12514	
OY	61 GATCTGCTTGAAGCCTGGAGGTTGAGGCTACAGTAAGCCAAAGTCATGCCAGTATCTTC 120	
DB	12515 GAATGGCATGAAACCACAAGAGGACAGGCTTGCATGTAGCCGAGATCACACATTGCATCC 12574	
OY	121 AGCTTGCGCGAACAAGTGAGAACCGCTAACAAAAAAAATTATAAAAGA 174	
DB	12575 AGCCTGGGTGACAAAGTAGAGCTCCGTCAAATAAAAAAAAAAATTATAAATATA 12628	
RESULT 7		
ABAO7813		
ID	ABAO7813 standard; DNA; 32134 BP.	
XX		
AC	ABAO7813;	
XX		
DT	11-JAN-2002 (first entry)	
DE	Human ovarian and breast cancer associated polynucleotide SEQ ID NO 608.	
XX		
KW	Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;	
KW	antiallergic; hepatotropic; antidiabetic; antiinflammatory; anticancer;	
KW	vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;	
KW	cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;	
XX	neurological disease; infection; human; secreted protein; ds.	
XX		



OS Homo sapiens.  
XX  
PN W0200155325-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001MO-US001345.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225470P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.

PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249246P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249266P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251866P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-025678P.  
PA (HUMA-) HUMAN GENOME SCT INC.  
XX



XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-488786/53.  
XX New isolated ovarian and/or breast cancer related nucleic acids and  
PT polypeptides, useful for diagnosing, treating and/or preventing human  
PT diseases and disorders, particularly ovarian and/or breast cancer.  
XX  
PS Disclosure; SEQ ID NO 608; 577pp + Sequence listing; English.  
XX  
XX The invention relates to novel genes (ABA07454-ABA08224) and proteins  
CC (ABA10743-ABA10980) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in  
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,  
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune  
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic  
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,  
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)  
CC cardiovascular disorders such as myocardial ischaemia; (d) wound healing  
CC / (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)  
CC infectious diseases such as viral, bacterial, fungal and parasitic  
CC infections. Note: The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in electronic format directly  
CC from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences)  
XX  
SQ Sequence 32134 BP; 9171 A; 6268 C; 6931 G; 9764 T; 0 U; 0 Other;  
  
Query Match 31.3%; Score 94; DB 4; Length 32134;  
Best Local Similarity 71.3%; Pred. No. 5.5e-13;  
Matches 124; Conservative 0; Mismatches 50; Indels 0; Gaps 0;  
  
QY 1 GGGGGCCCTGTAATCCAGCTACTCAGAGGCTGAGGAGGAGATCCGCGAGCTGGCA 60  
Db 12455 GGGCACTGTACTCTCCAGCTACTCGGAGGCTGAGGAGGAGATGAGGAGGAGCA 12514  
QY 61 GATCTGCTGAGCTGAGCTGAGGAGGCTACGTACGTACGATCATGCCATATCTTC 120  
Db 12515 GATGCGCATGATGAGGAGGAGGAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAG 12574  
QY 121 AGCTTGCGGCGCAAAAGTGAAGCCGTAAACAAAAAATTTAAAAAGAA 174  
Db 12575 AGCTTGCGGCGCAAAAGTGAAGCCGTCTCAAAAAAATTTAAAAATAA 12628  
  
RESULT 8  
AAL03616  
ID AAL03616 standard; DNA; 32191 BP.  
XX  
AC AAL03616;  
XX  
DT 21-NOV-2001 (first entry)  
XX  
DE Human reproductive system related antigen DNA SEQ ID NO: 6304.  
XX  
KM Human; reproductive system related antigen; reproductive system disorder;  
KM cancer; gene therapy; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200155320-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US001339.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.

PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 11-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217495P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225477P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226682P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227099P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0228287P.  
PR 01-SEP-2000; 2000US-0228343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0232402P.  
PR 14-SEP-2000; 2000US-0232403P.  
PR 14-SEP-2000; 2000US-0232404P.  
PR 14-SEP-2000; 2000US-0232405P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 21-SEP-2000; 2000US-0234275P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0234999P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235835P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236328P.  
PR 29-SEP-2000; 2000US-0236357P.  
PR 29-SEP-2000; 2000US-0236358P.  
PR 29-SEP-2000; 2000US-0236359P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.



PR	13-OCT-2000;	2000US-0239935P.
PR	13-OCT-2000;	2000US-0239937P.
PR	20-OCT-2000;	2000US-0240960P.
PR	20-OCT-2000;	2000US-0241221P.
PR	20-OCT-2000;	2000US-0241785P.
PR	20-OCT-2000;	2000US-0241786P.
PR	20-OCT-2000;	2000US-0241787P.
PR	20-OCT-2000;	2000US-0241808P.
PR	20-OCT-2000;	2000US-0241809P.
PR	20-OCT-2000;	2000US-0241826P.
PR	01-NOV-2000;	2000US-0244617P.
PR	08-NOV-2000;	2000US-0246474P.
PR	08-NOV-2000;	2000US-0246475P.
PR	08-NOV-2000;	2000US-0246476P.
PR	08-NOV-2000;	2000US-0246477P.
PR	08-NOV-2000;	2000US-0246478P.
PR	08-NOV-2000;	2000US-0246523P.
PR	08-NOV-2000;	2000US-0246524P.
PR	08-NOV-2000;	2000US-0246525P.
PR	08-NOV-2000;	2000US-0246526P.
PR	08-NOV-2000;	2000US-0246527P.
PR	08-NOV-2000;	2000US-0246528P.
PR	08-NOV-2000;	2000US-0246532P.
PR	08-NOV-2000;	2000US-0246602P.
PR	08-NOV-2000;	2000US-0246609P.
PR	08-NOV-2000;	2000US-0246610P.
PR	08-NOV-2000;	2000US-0246611P.
PR	17-NOV-2000;	2000US-0249211P.
PR	17-NOV-2000;	2000US-0249212P.
PR	17-NOV-2000;	2000US-0249213P.
PR	17-NOV-2000;	2000US-0249214P.
PR	17-NOV-2000;	2000US-0249215P.
PR	17-NOV-2000;	2000US-0249216P.
PR	17-NOV-2000;	2000US-0249217P.
PR	17-NOV-2000;	2000US-0249218P.
PR	17-NOV-2000;	2000US-0249244P.
PR	17-NOV-2000;	2000US-0249245P.
PR	17-NOV-2000;	2000US-0249246P.
PR	17-NOV-2000;	2000US-0249247P.
PR	17-NOV-2000;	2000US-0249265P.
PR	17-NOV-2000;	2000US-0249297P.
PR	17-NOV-2000;	2000US-0249299P.
PR	17-NOV-2000;	2000US-0249300P.
PR	01-DEC-2000;	2000US-0250160P.
PR	01-DEC-2000;	2000US-0250391P.
PR	05-DEC-2000;	2000US-0251030P.
PR	05-DEC-2000;	2000US-0251988P.
PR	05-DEC-2000;	2000US-0256719P.
PR	06-DEC-2000;	2000US-0251479P.
PR	08-DEC-2000;	2000US-0251856P.
PR	08-DEC-2000;	2000US-0251868P.
PR	08-DEC-2000;	2000US-0251869P.
PR	08-DEC-2000;	2000US-0251989P.
PR	11-DEC-2000;	2000US-0251990P.
PR	05-JAN-2001;	2001US-0255678P.

PA (HUMA-, HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-465570/50.  
XX  
XX  
XX Isolated nucleic acid molecule encoding a reproductive system antigen  
PT used in preventing, treating or ameliorating a medical condition.  
XX  
XX  
XX Disclosure; SEQ ID NO 6304; 1297bp + Sequence Listing; English.  
XX  
XX  
XX The present invention provides the protein and coding sequences of a  
CC number of human reproductive system related antigens. These can be used

CC in the prevention and treatment of reproductive system disorders,  
CC including cancer. The present sequence is a genomic sequence encoding a  
CC protein of the invention  
XX  
SQ Sequence 32191 BP; 9130 A; 6394 C; 6829 G; 9838 T; 0 U; 0 Other;

Query Match	31.3%	Score 94;	DB 4;	Length 32191;
Best Local Similarity	71.3%	Pred. No. 5.5e-13;		
Matches 124;	Conservative	0;	Mismatches 50;	Indels 0;
			Gaps	0;

Qy 1 GGGAGGCGCTGATATCCCACTACTCAGAGAGGTAGAGAGGATCCGGAGAGCTGGCA 60

Db 9503 GGGCAACCTGTACTCCCACTACTCGGAGAGCTGAGGCGAGATGGCATGAGGCGAGA 9567

Qy 61 GATCTGCTGAGCTCGGAGAGGTGAGGCTACAGTAAGCCAAAGATCATGCGAGTATATCTC 120

Db 9563 GAATGGCATGAACCCAGAGGCGAGGCTTGCAGTGAGCCGAGATCACACCATTTGCATCC 9622

Qy 121 AGCGTGGGCGCAAAAGTGAGACCGTAACTAAAAAATTTAAAAAAGCA 174

Db 9623 AGCGTGGGCGCAAAAGTGAGACTCCGCTCTCAAAAAAATTTAAAAATAA 9676

RESULT 9  
ABA07814  
ID ABA07814 standard; DNA; 32191 BP.

DT 11-JAN-2002 (first entry)

Human ovarian and breast cancer associated polynucleotide SEQ ID NO 609.

Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral; antiallergic; hepatotropic; antidiabetic; antiinflammatory; anticancer; vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic; cardiact; gene therapy; cancer; immune disorder; cardiovascular disorder; neurological disease; infection; human; secreted protein; da-

**Homo sapiens.**

W0200155325-A2

02-AUG-2001

17-JAN-2001; 2001WO-US001345.

31-JAN-2000; 2000US-0179065P

24-FEB-2000; 2000US-0184664P

16-MAR-2000; 2000US-0189874P

18-APR-2000; 2000US-0198123P

07-JUN-2000; 2000US-0209467P

30-JUN-2000; 2000US-0215135P  
07-JUL-2000; 2000US-0216647P

07-JUL-2000; 2000US-0216B80P  
11-TH-2000; 2000TE-0217497P

11-JUL-2000; 2000US-0217496P  
14-JUL-2000; 2000US-0218290P

26-JUL-2000; 2000TTS=0220963P  
36-JUL-2000; 2000TTS=0220964P

14-AUG-2000; 2000UTS-0224518P  
14-AUG-2000; 2000UTS-0224519P

14-AUG-2000; 2000UTS-0225213P  
14-AUG-2000; 2000UTS-0225214P

14-AUG-2000; 2000US-0225267P

14-AUG-2000; 2000US-0225268E  
14-AUG-2000; 2000US-0225279P



PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 23-AUG-2000; 2000US-0227182P.  
PR 30-AUG-2000; 2000US-02277009P.  
PR 01-SEP-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232337P.  
PR 14-SEP-2000; 2000US-0232339P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.

PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249246P.  
PR 17-NOV-2000; 2000US-0249254P.  
PR 17-NOV-2000; 2000US-0249255P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 05-DEC-2000; 2000US-0250130P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251855P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251893P.  
PR 08-DEC-2000; 2000US-0251909P.  
PR 11-DEC-2000; 2000US-0256097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
PI Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-488786/53.  
XX  
XX New isolated ovarian and/or breast cancer related nucleic acid and  
PT polypeptides, useful for diagnosing, treating and/or preventing human  
XX diseases and disorders, particularly ovarian and/or breast cancer.  
XX  
XX Disclosure; SEQ ID NO 609; 577bp + Sequence Listing; English.  
XX  
XX  
CC The invention relates to novel genes (ABA07454-ABA08224) and proteins  
CC (ABH10743-ABH10980) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in  
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,  
CC breast, gastrointestinal tract, liver, lung or urogenital; (b) immune  
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic  
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,  
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)  
CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing  
CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)  
CC infectious diseases such as viral, bacterial, fungal and parasitic  
CC infections. Note: The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in electronic format directly  
XX from WIFO at ftp.wifo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 32191 BP; 9130 A; 6394 C; 6829 G; 9838 T; 0 U; 0 Other;

Query Match 31.3%; Score 94; DB 4; Length 32191;  
Best Local Similarity 71.3%; Pred. No. 5.5e-13;  
Matches 124; Conservative 0; Mismatches 50; Indels 0; Gaps 0;



QY 1 GGGGGGCTGTATCCAGCTACTCAGAGGCTGAGGAGGATCCGGAGCTTGCA 60  
DB 9503 GGGGACCTTGTACTCCAGCTACTCAGAGGCTGAGGAGGATCCAGAGGAGCA 9562  
QY 61 GATCTGCTGAGGCTGAGGAGGTTGAGGCTACAGTAAGCCAGATCATGCACTTTC 120  
DB 9563 GAATGGATGACCCAGAGGAGGAGGCTTGCAGTGAAGGATGACCATTCGACTCC 9622  
QY 121 AGCCTGGGCGACAAAGTGAGACCGTAAACAAAAAATTTAAAAAGAA 174  
DB 9623 AGCCTGGGCGACAAAGTGAGACCGCTCAAAAAAATTTAAAAATATAA 9676  
RESULT 10  
ID ABR83570 standard; cDNA; 148834 BP.  
XX ABR83570;  
AC ABR83570;  
XX  
DT 14-AUG-2002 (first entry)  
XX  
DE Human cDNA differentially expressed in granulocytic cells #141.  
XX  
KW Human; ss; granulocytic cell; DNA chip; bacterial infection;  
KW viral infection; parasitic infection; protozoal infection;  
KW fungal infection; sterile inflammatory disease; psoriasis;  
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;  
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;  
KW adult respiratory distress syndrome; inflammatory bowel disease;  
KW Crohn's disease; ulcerative colitis; periodontal disease;  
KW granulocyte activation; chronic inflammation; allergy.  
XX  
XX Homo sapiens.  
OS  
XX  
FN W0200228999-A2.  
PD 11-APR-2002.  
XX  
PF 03-OCT-2001; 2001MO-US030821.  
XX  
PR 03-OCT-2000; 2000US-0237189P.  
XX  
XX (GENE-) GENE LOGIC INC.  
XX  
PI Beazer-Barclay Y, Weisman SM, Yamaga S, Vockley J;  
XX  
DR WPI; 2002-435328/46.  
XX  
XX  
PT Detecting granulocyte activation by detecting differential expression of  
PT genes associated with granulocyte activation, which serves as diagnostic  
PT markers that is useful for monitoring disease states and drug toxicity.  
XX  
PS Claim 1; SEQ ID NO 141, 114pp; English.  
XX  
XX The invention relates to detecting (M1) granulocyte (GC) activation  
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by  
CC DNA chip analysis as given in the specification, and comparing the  
CC expression level to an expression level in an unactivated GC where  
CC differential expression of Gs is indicative of GCA. Also included are  
CC modulating (M2) GA by contacting GC with an agent that alters the  
CC expression of at least one gene in Gs; (2) screening (M3) for an agent  
CC capable of modulating GCA or an inflammation (especially chronic) in a  
CC tissue, an allergic response in a subject, exposure of a subject to a  
CC pathogen or sterile inflammatory disease using the gene expression  
CC profile; (3) detecting (M4) an inflammation (especially chronic) in a  
CC tissue, an allergic response in a subject, exposure of a subject to a  
CC pathogen or sterile inflammatory disease, by detecting the level of  
CC expression in a sample of the tissue of gene(s) from Gs, where the level  
CC of expression of the gene is indicative of inflammation; (4) treating  
CC (M5) an inflammation (especially chronic) or in a tissue, an allergic  
CC response in a subject, exposure of a subject to a pathogen or sterile  
CC inflammatory disease, by contacting a tissue having inflammation with an

CC agent that modulates the expression of gene(s) from Gs in the tissue. M1  
CC is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful  
CC for screening an agent capable of modulating GCA preferably in an  
CC inflammation in a tissue; M4 is useful for detecting an inflammation  
CC (especially chronic) in a tissue, an allergic response in a subject,  
CC exposure of a subject to a pathogen or sterile inflammatory disease (e.g.  
CC psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis,  
CC cardiac reperfusion injury, renal reperfusion injury, ARDS, adult  
CC respiratory distress syndrome, inflammatory bowel disease, Crohn's  
CC disease, ulcerative colitis, periodontal disease, also bacterial  
CC infection, viral infection, parasitic infection, protozoal infection,  
CC fungal infection and M5 is useful for treating one of the above  
CC conditions. The present sequence represents a gene differentially  
CC expressed in granulocytes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 148834 BP; 38942 A; 33672 C; 34501 G; 41719 T; 0 U; 0 Other;  
XX  
Query Match 31.3%; Score 93.8; DB 6; Length 148834;  
Best Local Similarity 70.6%; Pred. No. 8.9e-13;  
Matches 125; Conservative 0; Mismatches 52; Indels 0; Gaps 0;  
QY 1 GGGGGGCTGTATCCAGCTACTCAGAGGCTGAGGAGGATCCGGAGCTTGCA 60  
DB 86336 GGGGCTTGTACTCCAGCTACTCAGAGGCTGAGGAGGATCCAGAGGAGCA 86395  
QY 61 GATCTGCTGAGGCTGAGGAGGTTGAGGCTACAGTAAGCCAGATCATGCACTTTC 120  
DB 86336 GAATGGATGACCCAGAGGAGGAGGCTTGCAGTGAAGGATGACCATTCGACTCC 86455  
QY 121 AGCCTGGGCGACAAAGTGAGACCGTAAACAAAAAATTTAAAAAGAAAT 177  
DB 86456 AGCCTGGGCGACAAAGTGAGACCGCTCAAAAAAATTTAAAAATATAA 86512  
RESULT 11  
ABT34215/C  
ID ABT34215 standard; DNA; 24420 BP.  
XX  
XX ABT34215;  
AC  
XX  
DT 12-JUN-2003 (first entry)  
XX  
DE Human serotonin receptor 1D gene SEQ ID No 1.  
XX  
KW Eating disorder; polymorphism; dataset; allele; HGBASE identification;  
KW serotonin receptor 1D; delta-opioid receptor; dopamine receptor D2;  
KW anorexia nervosa; bulimia nervosa; human; ds.  
XX  
OS Homo sapiens.  
XX  
XX W02003012143-A1.  
PN  
XX  
PD 13-FEB-2003.  
XX  
PF 16-JUL-2002; 2002MO-US022555.  
XX  
PR 16-JUL-2001; 2001US-0305133P.  
XX  
PR 20-NOV-2001; 2001US-0306440P.  
XX  
PR 13-NOV-2001; 2001US-0331285P.  
XX  
PR 19-DEC-2001; 2001US-0340843P.  
XX  
PR 19-DEC-2001; 2001US-0340844P.  
XX  
PA (PRIC-) PRICE FOUND LTD.  
XX  
PI Bergen AW, Yeager M;  
XX  
XX WPI; 2003-268122/26.  
DR  
XX  
PT New nucleic acid molecule having polymorphisms in the serotonin receptor  
PT ID, delta-opioid receptor, or dopamine receptor D2, useful in diagnostic



PT and prognostic assays for eating disorders, such as anorexia and bulimia nervosa.

XX Disclosure; Page 94-107; 149pp; English.

PS The invention relates to a novel isolated nucleic acid molecule

CC comprising a variant gene associated with an eating disorder and selected

CC from any of 119 polymorphisms with their corresponding genotyping in

CC dataset, alleles and HBBASE identification, given in the specification.

CC The novel nucleic acid molecule has polymorphisms in the serotonin

CC receptor 1D, delta-opioid receptor, or dopamine receptor D2, which is

CC useful in diagnostic and prognostic assays for eating disorders, in

CC particular anorexia nervosa and bulimia nervosa. This polynucleotide

CC sequence represents a human serotonin receptor 1D gene of the invention

XX

SQ Sequence 24420 BP; 6212 A; 5915 C; 5886 G; 6407 T; 0 U; 0 Other;

Query Match 31.2%; Score 93.6; DB 7; Length 24420;

Best Local Similarity 71.5%; Pred. No. 6.4e-11;

Matches 123; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 3 GGGCGCTGTAATCCAGCTACTCAGAGGCTGAGCAGAGATCCGCGAGCTTGGCAGA 62

Db 17317 GTGCCTACAGTCCAGCTACTCGGAGGCTGAAGTGGAGAGCACCCTGATCTGGAGG 17258

QY 63 TCTGCTGAGGCTTGGAGGCTTGAAGCTAAGTAAAGCAATCATGCTGATTAATTCTGAG 122

Db 17257 TCCACCTGAGTCTGGAGGCTCAGGCGCGAGTGAAGTGAATGTCACCTGCACTCCAG 17198

QY 123 CCTGGCGCAAAAGTGAGACCGCTAACAAAAAAATTTTAAAAAAGA 174

Db 17197 CCTAGGTGACAGACAGTACCCCTGCTCAAAAAAGAAAAAGAAACAAAGAAAA 17146

RESULT 12

AAC93353

ID AAC93353 standard; cDNA; 1877 BP.

XX AAC93353;

XX

DT 16-FEB-2001 (first entry)

XX

DE Human secreted protein cDNA sequence #44.

XX

KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;

KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;

KW valiantry; anticonvulsant; antibacterial; antifungal; antiparasitic;

KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;

KW neurological disease; infection; human; secreted protein; ss.

XX

OS Homo sapiens.

XX

PN WO200058495-A1.

XX

PD 05-OCT-2000.

XX

PF 23-MAR-2000; 2000WO-US007661.

XX

PR 26-MAR-1999; 99US-0126504P.

XX

PR 07-JAN-2000; 2000US-0174847P.

XX

PA (HOMA-) HUMAN GENOME SCI INC.

XX

PI Rosen CA, Ruben SM, Komatsoulis G;

XX

DR WPI; 2000-611720/58.

XX

DR P-PSDB; AAB51422.

XX

XX New nucleic acid molecules encoding 45 human secreted proteins for

PT diagnosing, preventing, treating or ameliorating medical conditions and

PT used as food additives or preservatives.

XX

PS Claim 1; Page 352; 410pp; English.

XX

CC The invention relate to the isolation of genes AAC93310-C93354 encoding

CC 45 human secreted proteins AAB51380-B51423. The genes can be used to

CC generate fusion proteins by linking to the gene for the human

CC immunoglobulin G Fc portion (AAC93301) for increasing the stability of

CC the fusion protein as compared to the human protein only. The genes and

CC proteins are useful for preventing, ameliorating or treating medical

CC conditions, e.g. by protein or gene therapy. The genes are isolated from

CC a range of human tissues disclosed in the specification. The nucleic

CC acids, proteins, antibodies and (ant)agonists are useful in the

CC diagnosis, treatment and prevention of: (a) cancer, e.g. breast and

CC ovarian cancer, and other cancers of the adrenal gland, bone, bone

CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b)

CC immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic

CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,

CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)

CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing

CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)

CC infectious diseases such as viral, bacterial, fungal and parasitic

XX

SQ Sequence 1877 BP; 543 A; 358 C; 372 G; 596 T; 0 U; 8 Other;

Query Match 31.1%; Score 93.4; DB 3; Length 1877;

Best Local Similarity 72.5%; Pred. No. 3.8e-13;

Matches 121; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 3 GGGCGCTGTAATCCAGCTACTCAGAGGCTGAGCAGAGATCCGCGAGCTTGGCAGA 62

Db 1706 GGGCGCTGTAATCCAGCTACTCAGAGGCTGAGCAGAGATCCGCGAGCTTGGCAGA 1765

QY 63 TCTGCTGAGGCTTGGAGGCTTGAAGCTAAGTAAAGCAATCATGCTGATTAATTCTGAG 122

Db 1766 ATTGCTTGAACCCGGAGGCGGAGGCTGCAATGAGCCAGATCATGCTGATGCACTCCAG 1825

QY 123 CCTGGCGCAAAAGTGAGACCGCTAACAAAAAAATTTTAAAA 169

Db 1826 TCTAGGCAACAGGCGCAAAATCCATCTCCAAAAAAATTTTAAAA 1872

RESULT 13

AAL38336/c

ID AAL38336 standard; DNA; 143899 BP.

XX AAL38336;

XX

DT 15-AUG-2002 (first entry)

XX

DE Genomic sequence encoding a human NGR2 protein.

XX

OS Homo sapiens.

XX

PN WO200229059-A2.

XX

PD 11-APR-2002.

XX

PF 06-OCT-2001; 2001WO-US031488.

XX

PR 06-OCT-2000; 2000US-0238361P.

XX

XX (UYA) UNIV YALE.

XX

PA (BIO) BIOGEN INC.

XX

XX Cerebroprotective; neuroprotective; cytosstatic; Nogo receptor homologue;

KW NGR2; NGR3; axonal growth; central nervous system; CNS; cerebral injury;

KW spinal cord injury; stroke; demyelinating disease; multiple sclerosis;

KW monophasic demyelination; encephalomyelitis; Marchiafava-Bignam disease;

KW multifocal leukoencephalopathy; panencephalitis; Spongy degeneration;

KW Alexander's disease; Canavan's disease; metachromatic leukodystrophy;

KW Krabbe's disease; immune; bait protein; genetic mapping; gene therapy;

KW transgenic animal; unregulated cellular growth; cancer; tumour; human;

XX

gene; ds.

XX

OS Homo sapiens.

XX

PN WO200229059-A2.

XX

PD 11-APR-2002.

XX

PF 06-OCT-2001; 2001WO-US031488.

XX

PR 06-OCT-2000; 2000US-0238361P.

XX

XX (UYA) UNIV YALE.

XX

PA (BIO) BIOGEN INC.

XX



PI Scitlatter SM, Cate RJ, Sah DWY;  
 DR WPI; 2002-416677/44.  
 XX  
 XX Novel Nogo receptor homolog polypeptide, NGR2 or NGR3, useful for  
 PT treating central nervous system disorder, cerebral injury, spinal cord  
 PT injury, stroke, and demyelinating diseases.  
 XX  
 XX Example 2; Page 176-214; 277pp; English.

XX The invention relates to a Nogo receptor homologue polypeptide, NGR2 or  
 CC NGR3, comprising a 50 amino acid LRRT sequence, a 284 amino acid NTLRRT  
 CC sequence, or a 420, 461 or 392 amino acid sequence, all given in the  
 CC specification. The NGR3 protein or its binding antibody is useful for  
 CC decreasing inhibition of axonal growth of a central nervous system (CNS)  
 CC neuron, by contacting the neuron NGR3 or its antibody, and for treating  
 CC CNS disease, disorder or injury. NGR3 or a vector comprising NGR3 is  
 CC useful for treating cerebral injury, spinal cord injury, stroke,  
 CC demyelinating diseases, e.g. multiple sclerosis, monophasic  
 CC demyelination, encephalomyelitis, multifocal leukoencephalopathy,  
 CC paraneoplastic, Marchiafava-Bignami disease, spongy degeneration,  
 CC Alexander's disease, Canavan's disease, metachromatic leukodystrophy and  
 CC Krabbe's disease. NGR3 is useful for inducing an immune response in a  
 CC mammal against NGR3, as a bait protein in a two-hybrid or three-hybrid  
 CC assay, and as a research tool for identification, characterization and  
 CC purification of interacting, regulatory proteins. The nucleotide  
 CC sequences of the invention are useful for screening for RFP associated  
 CC with certain disorders, for genetic mapping, and for gene therapy. The  
 CC vector containing NGR3 is useful for producing non-human transgenic  
 CC animals. The NGR3 binding antibody is useful for isolating and purifying  
 CC NGR3, for localisation and/or quantitation of NGR3, and for diagnostic  
 CC and therapeutic purposes. The sequences of the invention, vectors and  
 CC antibodies are useful for treating or preventing unregulated cellular  
 CC growth such as cancer and tumour growth. This polynucleotide sequence  
 CC represents the genomic sequence encoding a human NGR2 protein of the  
 CC invention

SO Sequence 143899 BP; 36346 A; 35277 C; 35318 G; 35657 T; 0 U; 1301 Other;

Query Match 31.1%; Score 93.2; DB 6; Length 143899;  
 Best Local Similarity 71.8%; Pred. No. 1.2e-12;  
 Matches 122; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY CCTGTAATCCAGCTACTCAGAGGCTGAGGACAGAGAGATCCGGGAGCTGGCAATCT 65  
 Db CCTTAACTTCTAGCTACTTGGAGGCTGAGGATCACTTGGACATGGGAGGTC 3876  
 QY GCCTGAGGCTGGAGGTTGAGGCTACAGTAAGCAAGATCATGCGAGTAATCTCAGCT 125  
 Db ACTTGAGCTTGGAGGTTGAGGCTGCGCATGAGCCATGATCAACCACTGCATCTCCATCT 3816  
 QY 126 GGGCGACAAAGTGAAGCCGTAAACAAAAAATTTTAAAAAAGAAA 175  
 Db 3815 GGGTGACAGATGAGATGTCTTAAATAAATTTTAAAAAATCAAAA 3766

RESULT 14  
 AAC04760/c  
 ID AAC04760 standard; cDNA; 203 BP.

AC AAC04760;  
 XX  
 XX 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 8835.

KM Human, 5' EST, expressed sequence tag, secreted protein; cDNA isolation;  
 KM gene therapy; chromosome mapping; ss.

OS Homo sapiens.

XX Homo sapiens.  
 XX EP1033401-A2.  
 XX

PD 06-SEP-2000.

XX 21-FEB-2000; 2000EP-00200610.

PR 26-FEB-1999; 99US-0122487P.

PA (BEST) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures.  
 XX  
 XX Claim 1; SEQ ID NO 8835; 71pp + Sequence listing; English.

XX The present sequence is one of a large number of 5' ESTs derived from  
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively  
 CC identified within the present sequence. The 5' ESTs were prepared from  
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST  
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)  
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA  
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences  
 CC derived from the 5' ends of mRNAs and even in those cases where longer  
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'  
 CC ESTs are derived from mRNAs with intact 5' ends and can therefore be used  
 CC to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in  
 CC diagnostic, forensic, gene therapy and chromosome mapping procedures.  
 CC They are used to obtain upstream regulatory sequences and to design  
 CC expression and secretion vectors

SO Sequence 203 BP; 29 A; 60 C; 42 G; 66 T; 0 U; 6 Other;

Query Match 30.7%; Score 92; DB 3; Length 203;  
 Best Local Similarity 67.2%; Pred. No. 4.9e-13;  
 Matches 121; Conservative 5; Mismatches 53; Indels 0; Gaps 0;

QY 2 GGGGCTGTAATCCAGCTACTCAGAGGCTGAGGACAGAGAGATCCGGGAGCTGGCAG 61  
 Db 183 GGTGCTGTAATCCAGCTACTCAGAGGCTGAGGACAGAGAGATCCGGGAGCTGGCAG 124  
 QY 62 ATCTGCTGAGGCTGGAGGTTGAGGCTACAGTAAGCAAGATCATGCGAGTAATCTCA 121  
 Db 122 AATGCTGTAATCCAGGAGGTTGAGGCTTGCAGTGAAGCAATGGACATCACTCAATCCA 64  
 QY 122 GCCTGGGCGACAAAGTGAAGCCGTAAACAAAAAATTTTAAAAAAGAAA 181  
 Db 63 GCCTGGGCGACAGACGAGATCCCTTAATAAATAAATAAATAAATAAATAAATAAATA 4

RESULT 15  
 ABS69895/c  
 ID ABS69895 standard; DNA; 56737 BP.

AC ABS69895;  
 XX  
 XX 21-NOV-2002 (first entry)

DE Human hypoxanthine-guanine phosphoribosyltransferase (HPRT) gene.

KM Vector; adenovirus; adeno-associated; adenosine deaminase gene; receptor;  
 KM adenosine deaminase deficiency; severe combined immune deficiency; PAH;  
 KM beta-chain; haemoglobin gene; beta-thalassemia; sickle cell disease;  
 KM low density lipoprotein gene; familial hypercholesterolaemia;  
 KM hypoxanthine-guanine phosphoribosyltransferase; Lesch-Nyhan syndrome;  
 KM phenylalanine hydroxylase gene; gene therapy; phenylketonuria;  
 KM dystrophin gene; muscular dystrophy; cystic fibrosis; immunosuppressant;  
 KM human cystic fibrosis transmembrane conductance regulator gene;  
 KM antihaemic; antilipemic; nootropic; cytosaric; dermatological; human;  
 KM alpha-1-antitrypsin; lysosomal glucocerebrosidase; ADA; HPRT;  
 KM lysosomal arylsulphatase A; ornithine transcarbamylase; ARSA; OTC; NF;



KM purin nucleoside phosphorylase; gene; ds.

OS Homo sapiens.

PN US2002102731-A1.

PD 01-AUG-2002.

PF 12-FEB-2001; 2001US-00782378.

PR 02-OCT-2000; 2000US-0237747P.

PA (UYNV ) UNIV NEW YORK STATE RES FOUND.

PI Hearing P, Bahou WF, Sandalon Z, Gatenko DV;

DR WPI; 2002-690619/74.

PT Producing vector, by introducing vector having nucleotide sequence,  
PT adenovirus inverted terminal repeats and packaging sequence, and adeno-  
PT associated virus terminal repeat, into cell, and culturing cell.

PS Disclosure; Page 93-117; 191pp; English.

CC The present invention relates to a new method of producing a vector. The  
CC method involves introducing recombinant vector having nucleotide sequence  
CC (NS) having 5' and 3' end, left and right inverted terminal repeats of  
CC adenovirus flanking NS, adenovirus packaging sequence linked to inverted  
CC terminal repeat, and adeno-associated virus terminal repeat linked to 3'  
CC end of NS, into cell expressing adenovirus early gene lacking from vector  
CC ; and culturing cell to produce another vector. The method is useful for  
CC generating vectors, especially mad vectors. The method is useful in  
CC transferring nucleotide sequences of interest into a cell, for gene  
CC transfer applications (e.g. gene therapy) in vitro, ex vivo and in vivo.  
CC The nucleotide sequences are useful for treating diseases associated with  
CC 1c, i.e. adenosine deaminase gene associated with adenosine deaminase  
CC deficiency with severe combined immune deficiency, beta-chain of  
CC haemoglobin gene associated with beta-thalassemia and sickle cell  
CC disease, receptor for low density lipoprotein gene associated with  
CC familial hypercholesterolemia, hypoxanthine-guanine  
CC phosphoribosyltransferase associated with Lesch-Nyhan syndrome,  
CC phenylalanine hydroxylase (PAH) gene associated with phenylketonuria,  
CC dystrophin gene associated with muscular dystrophy, and human cystic  
CC fibrosis transmembrane conductance regulator gene associated with cystic  
CC fibrosis. The present nucleic acid sequence represents a human disease  
CC gene sequence that was used in the methods of the invention

SQ Sequence 56737 BP; 15689 A; 11281 C; 11599 G; 18168 T; 0 U; 0 Other;

Query Match 30.3%; Score 91; DB 6; Length 56737;

Best Local Similarity 65.9%; Pred. No. 3,4e-12; Matches 147; Conservative 0; Mismatches 75; Indels 1; Gaps 1;

QY 1 GGGGGCCTGTAATCCAGCTACTCAAGAGGCTGAGGCGAGGATCCGCGAGCTGCG- 59

DB 3504 GCGCACTGTAGTCCGAGCTACTTGGAGGCTGAAGTGGAGAAATCCCTGAGCTGGGG 3445

QY 60 AGATTCGCTGAGCTGTGGAGGTTGAGGCTACAGTAAGCCAAAGATCATGCCAGTATCTT 119

DB 3444 AGATCAACCCGAGCCCGGAGAGTCAAGGCTGAGTGAGCGAGTGTGTGCTGCTGCACCTC 3385

QY 120 CAGGCTGGGAGCAAGAGTGAACCGGTACAAAAAATTTAAAAAAGAAATTTA 179

DB 3384 CATCTAGGTGACGAGATGAGACCCCTGTCTCAAAAAAAGAAATTTGCGCAATTTAACTAA 3325

QY 180 GATCAAGATCCAACTGTAAAAAGTGGCTTAAACACCAATTA 222

DB 3324 GTTCATGTTTGAAGATGAAAAATCAACATTTTTCCTCAGCAA 3282



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 12, 2004, 17:39:25 / Search time 22.3214 Seconds  
(without alignments)  
7458.540 Million cell updates/sec

Title: US-09-966-880A-10\_COPY\_1\_300

Perfect score: 300

Sequence: 1 9999GCGCTGATCCAGCT.....AATGCGTGGTGCACCTGCT 300

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA: \*  
1: /cgm2\_6/ptodata/2/ina/5A.COMB.seq: \*  
2: /cgm2\_6/ptodata/2/ina/5B.COMB.seq: \*  
3: /cgm2\_6/ptodata/2/ina/6A.COMB.seq: \*  
4: /cgm2\_6/ptodata/2/ina/6B.COMB.seq: \*  
5: /cgm2\_6/ptodata/2/ina/PTUS.COMB.seq: \*  
6: /cgm2\_6/ptodata/2/ina/backfile1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	82.8	27.6	63000	4 US-09-780-172-18	Sequence 18, Appl
C 2	81.6	27.2	48763	4 US-09-916-204-3	Sequence 3, Appl
C 3	81.4	27.1	38059	4 US-09-328-925-4	Sequence 4, Appl
C 4	81.2	27.1	168575	4 US-09-426-290-1	Sequence 1, Appl
C 5	81	27.0	66804	4 US-09-740-041-3	Sequence 3, Appl
C 6	80.8	26.9	1001	4 US-09-671-317-12	Sequence 12, Appl
C 7	80.8	26.9	87350	3 US-08-781-891-79	Sequence 79, Appl
C 8	80.8	26.9	87350	4 US-09-618-166-79	Sequence 79, Appl
C 9	80.8	26.9	87543	4 US-09-791-211-3	Sequence 3, Appl
C 10	80.8	26.9	319608	4 US-09-539-333D-1	Sequence 1, Appl
C 11	80.8	26.9	319608	4 US-09-678-409-1	Sequence 1, Appl
C 12	80.6	26.8	7130	3 US-09-056-105-31	Sequence 1, Appl
C 13	80.4	26.8	367	4 US-09-702-705-1626	Sequence 1626, Ap
C 14	80.4	26.8	367	4 US-09-736-457-1626	Sequence 1626, Ap
C 15	80.4	26.8	367	4 US-09-614-124B-1626	Sequence 1626, Ap
C 16	80.4	26.8	367	4 US-09-671-325-1626	Sequence 1626, Ap
C 17	80.2	26.7	55298	4 US-09-491-356C-1	Sequence 1, Appl
C 18	80	26.7	42571	4 US-09-810-347-3	Sequence 3, Appl
C 19	79.6	26.5	2841	4 US-09-526-193A-24	Sequence 24, Appl
C 20	79.4	26.5	1624	2 US-08-852-807-10	Sequence 10, Appl
C 21	79.4	26.5	5835	3 US-09-033-333-3	Sequence 3, Appl
C 22	79.4	26.5	5835	4 US-09-033-556-2	Sequence 2, Appl
C 23	79.4	26.5	5835	4 US-09-614-495-3	Sequence 3, Appl
C 24	79.4	26.5	5835	4 US-09-151-376-2	Sequence 2, Appl
C 25	79.4	26.5	5836	1 US-08-380-916-1	Sequence 1, Appl
C 26	79.4	26.5	5836	1 US-08-721-690-1	Sequence 1, Appl
C 27	79.4	26.5	5836	3 US-08-891-581-1	Sequence 1, Appl

C 28	79.4	26.5	5836	3 US-09-033-333-2	Sequence 2, Appl
C 29	79.4	26.5	5836	4 US-09-033-556-1	Sequence 1, Appl
C 30	79.4	26.5	5836	4 US-09-614-495-2	Sequence 2, Appl
C 31	79.4	26.5	5836	4 US-09-474-692-9	Sequence 9, Appl
C 32	79.4	26.5	5836	4 US-09-151-376-1	Sequence 1, Appl
C 33	79.4	26.5	13674	2 US-08-852-807-1	Sequence 1, Appl
C 34	79.2	26.4	174493	4 US-09-804-471A-3	Sequence 3, Appl
C 35	79.2	26.4	174493	4 US-10-238-709-3	Sequence 3, Appl
C 36	79	26.3	12597	4 US-09-705-292-12	Sequence 12, Appl
C 37	79	26.3	116592	4 US-09-818-512-3	Sequence 3, Appl
C 38	78.8	26.3	308	4 US-09-702-705-1111	Sequence 1111, Ap
C 39	78.8	26.3	308	4 US-09-736-457-1111	Sequence 1111, Ap
C 40	78.8	26.3	308	4 US-09-614-124B-1111	Sequence 1111, Ap
C 41	78.8	26.3	308	4 US-09-671-325-1111	Sequence 1111, Ap
C 42	78.8	26.3	364	4 US-09-702-705-867	Sequence 867, App
C 43	78.8	26.3	364	4 US-09-736-457-867	Sequence 867, App
C 44	78.8	26.3	364	4 US-09-614-124B-867	Sequence 867, App
C 45	78.8	26.3	364	4 US-09-671-325-867	Sequence 867, App

## ALIGNMENTS

RESULT 1  
US-09-780-172-18/C  
; Sequence 18, Application US/09780172  
; Patent No. 6607916  
; GENERAL INFORMATION:  
; APPLICANT: Robert McKay  
; APPLICANT: Susan M. Preler  
; APPLICANT: Jacqueline Wyatt  
; TITLE OF INVENTION: ANTISENSE MODULATION OF CASEIN KINASE 2-ALPHA EXPRESSION  
; FILE REFERENCE: RTS-0159  
; CURRENT APPLICATION NUMBER: US/09/780,172  
; CURRENT FILING DATE: 2001-02-08  
; NUMBER OF SEQ ID NOS: 96  
; SEQ ID NO 18  
; LENGTH: 63000  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; US-09-780-172-18

Query Match 27.6%; Score 82.8; DB 4; Length 63000;  
Best Local Similarity 68.0%; Pred. No. 2.2e-12;  
Matches 140; Conservative 0; Mismatches 47; Indels 19; Gaps 1.

QY 1 GGGGCGCTGTATATCCAGCTACTCAGAGGCTGAGGCGAGAGATCCGCGAGCCTGACA 60  
DB 19400 GGGGCGCTGTATATCCAGCTACTCAGAGGCTGAGGCGAGAGATG----- 19355  
QY 61 GATTCGCCGAGCCTGGGAGGTTGAGGCTACAGTAAGCAAGATCATCCAGTACTTC 120  
DB 19354 -----GCGTGAACCCGAGAGGAGGCTTGAGGAGCGAGATGCTCCACTGCATCC 19300  
QY 121 AGCCTGGGCGAGCAAGTGAAGCCGTACACAAAAAATTTTAAAAAAGAAATTTAG 180  
DB 19299 AGCCTGGGCGAGAGCGAGAGCTCATCTCAAAAAAATTTTAAAAAATTTA 19240  
QY 181 ATCAAGATCAACTGTAAAAAGTGGC 206  
DB 19239 ATTAATATTAGGCTGTGCACAGTGGC 19214

RESULT 2  
US-09-916-204-3/C  
; Sequence 3, Application US/09916204  
; Patent No. 6618745  
; GENERAL INFORMATION:  
; APPLICANT: WEI, Ming-Hui et al.  
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
; TITLE OF INVENTION: THEREOF



```

1 /
2 / FILE REFERENCE: CLO001164CIP
3 / CURRENT APPLICATION NUMBER: US/09/916,204
4 / CURRENT FILING DATE: 2001-07-24
5 / NUMBER OF SEQ ID NOS: 6
6 / SOFTWARE: fastseq for Windows Version 4.0
7 /
8 / SEQ ID NO: 3
9 / LENGTH: 48763
10 /
11 / TYPE: DNA
12 /
13 / ORGANISM: Human
14 /
15 / US-09-916-204-3

```

Query Match	27.2%;	Score 81.6;	DB 4;	Length 48763;
Best Local Similarity	68.5%;	Pred. No. 4.1e-12;		
Matches 137;	Conservative 0;	Mismatches 44;	Indels 19;	Gaps 1

Cy	2GGGGCCCTGTAATCCAGCTACTCAGAGAGCTGAGGCGAGAGATCCGCGGAGCCTGGCAG	61
Dd	16610 GGCACCTTAATCCAGCTACTCAGAGAGCTAGGCGAGGAAT-----	16567

62 ATCTGCTCGAGCCTGGGAGGTTAGAGCTACAGTAAGCAAGATCATCCAGTAACTTCA 121  
 Oy  
 16566 ----TGTCTGAATCTCAGGAGGTGAGGTTGCAGTGAACCAAGATGGACCACTGCACATCCA 16510  
 Db

0y 122 GCTGGGCGCAAAAGTAGACCTTAACTAAAAAAGAAATTGGA 181  
Db 16509 GCTGGGTGACAGCAAGACTTACCTTAAAAAATTTAAAAA 16455

Qy	182	TCAGATCCAACTGTAAAA	201
Db	1649	GAATAAACCAAAAAAAAA	16430

RESULT 3  
US-09-328-925-4  
; Sequence 4, Application US/09328925

```

; APPLICANT: Kurauchi, Kotoku
; APPLICANT: Kurauchi, Sumiko
; TITLE OF INVENTION: Nucleotide Sequences for Gene Regulation and Methods of
; TITLE OF INVENTION: Use Thereof

```

```

; CURRENT APPLICATION NUMBER: US/09/328,923
; CURRENT FILING DATE: 1999-06-09
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO. 4

```

```

; LENGTH: 38059
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-328-925-4

```

Query Match	27.1%;	Score 81.4;	DB 4;	Length 38059;
Best Local Similarity	65.8%;	Pred. No. 4.3e-12;		
Matches 144;	Conservative	0;	Mismatches 56;	Indels 19;
				Gaps 1;

QY	GGGGGCTGTATCCAGCTACTACGAGAGGCTGAGCGAGGAGATCCGCGAGCTGGCA	60
Db	GGGTCCTGTATCCAGCTACTTGGAGGCTGAGCGAGGAGAT--	7502

QY 61 GATCGCCCTAGCGCTGGGAGGTTTGAGGCTAGTAGCCAGATCATGCCAGTATACCTC 120

Db 7503 -----GTTGACCTGGGAGATGGAGGTTGCACTGATGCTGATGCACCACTGCACCTCC 755

QY 121 AGCCGCGCCGCAAAAGTGAAGCCGTAAACAAAAAATTTAAAAAAGAAATTTAG 180  
Db 7558 AGCCTCGGCGAGAGCAAGCTCTGTCAAAAAAAGAGAGTGAAGAGAAAAGTCTAG 761

Dy 181 ATCAAGATCCAACTGTAAAAAGTGGCCTTAACACACCAT 219  
| | | | |  
Db 7618 GCTAATTCGAAGAATAAGTGAGCCCAAAGGAACTT 7656

RESULT 4  
US-09-426-290-1/c  
1. Sequence 1 Position 175/09436290

```

; Patent No. 6410712
;
; GENERAL INFORMATION:
;
; APPLICANT: Berglind Ran Olafsdottir
;
; APPLICANT: Jaffrey Gulcher

```

```

; TITLE OF INVENTION: HUMAN NARCOLEPSY GENE
; FILE REFERENCE: 2345.2001-000
; CURRENT APPLICATION NUMBER: US/09/426,290
; CURRENT FILING DATE: 1999-10-25

```

```
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
TGGCTGTTTG 168575
```

```

; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
NAME/KEY: CDS

```

```
; LOCATION: (21181) ... (21403)  
; NAME/KEY : CDS  
; LOCATION: (95252) ... (95430)  
; NAME/KEY : CDS
```

```

; LOCATION: (101753) ... (101996
; NAME/KEY: CDS
; LOCATION: (110324) ... (110439)
; NAME/KEY: CDS

```

```

; LOCATION: (124058) ... (124278
; NAME/KEY: CDS
; LOCATION: (127009) ... (127130
; NAME/KEY: CDS

```

US-09-426-290-1

Query Match	27.1%	Score 81.2	-DB 4	Length 168575
Best Local Similarity	71.4%	Pred. No. 8e-12	33	Indels 19
Matches 130	0	Mismatches	19	Gaps 1

Dy I GGGGCGCTGTATCCCAAGTACTCAGGAGGCTGAAGGCACGAGAGATCCGCCAGGCTGGCA 60  
|||  
Db 88612 GGGAGCTGTATCCCAAGTACTCAGGAGGCTGAAGGCACGAGAAATC----- 88

QY	61	GATCTGCGTGGAGGTTGAGGCTACAGTAAGCCAGATCATGCCAGTATACCTC	120
Db	88566	-----GCTTGAACCGGAGTGGAGGCTGAGTGAAGCCAGATTCACCACTGCACCTCC	88512

Qy 121 AGCTGGGCGACAAAGTAGAGCCGTACAAAAAAAAAAAAAAAAAATTAAATAAGAAATTTAG 180  
Db 88511 AGCTGGGCGACAGTAGAGCTCCCTTCACAAAAAAAAAAAAAAAAAAAAAAAAAATTAG 88452

QY	181	AT	182
Db	88451	TT	88450

RESULT 5  
US-09-740-041-3/c

; Patent No. 6562593  
 ; GENERAL INFORMATION:  
 ; APPLICANT: MERKULOV, Karl et al

```

1  TITLE OF INVENTION:  NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS
2
3  TITLE OF INVENTION:  AND USES THEREOF
4
5  FILE REFERENCE:  CLO01001
6

```

```

; CURRENT FILING DATE: 2000-12-20
;
; NUMBER OF SEQ ID NOS: 4
;
; SOFTWARE: FastSeq for Windows Version 4.0
;
; SEQ ID NO 3

```

```

; LENGTH: 66804
; TYPE: DNA
; ORGANISM: Human

```



US-09-740-041-3

## Query Match

27.0%; Score 81; DB 4; Length 66804;

Best Local Similarity 68.9%; Pred. No. 6.6e-12; Indels 0; Gaps 0;

## Matches 111; Conservative 0; Mismatches 50;

Qy 1 GGGGGCTGTATATCCAGTACTACAGAGGCTGAGGAGGATCCGGAGGAGCTGAGATCT 65

Db 35449 GTGTGCTGTATATCCAGTACTATTTAGAGGCTGAGGAGGATCCGGAGGAGCTGAGATCT 117

Qy 61 GATCTGCTGAGGCTGAGGAGGCTGAGGCTGAGGAGGATCCGGAGGAGCTGAGATCT 125

Db 35389 TCCTGCTGAGGCTGAGGAGGCTGAGGAGGATCCGGAGGAGCTGAGATCT 117

Qy 121 AGCTTGGGCGCAAGTGAAGCTTACCAAAAAA 161

Db 35329 AGCTTGGGCGCAAGTGAAGCTTACCAAAAAA 35289

## RESULT 6

US-09-671-317-12/c

Sequence 12, Application US/09671317

Patent No. 6528260

GENERAL INFORMATION:

APPLICANT: Blumenfeld, Marla

APPLICANT: Blumenfeld, Marla

APPLICANT: Blumenfeld, Marla

APPLICANT: Blumenfeld, Marla

APPLICANT: Blumenfeld, Marla

APPLICANT: Blumenfeld, Marla

APPLICANT: Blumenfeld, Marla

APPLICANT: Blumenfeld, Marla

APPLICANT: Blumenfeld, Marla

APPLICANT: Blumenfeld, Marla

APPLICANT: Blumenfeld, Marla

APPLICANT: Blumenfeld, Marla

APPLICANT: Blumenfeld, Marla

APPLICANT: Blumenfeld, Marla

APPLICANT: Blumenfeld, Marla

APPLICANT: Blumenfeld, Marla

APPLICANT: Blumenfeld, Marla

APPLICANT: Blumenfeld, Marla

APPLICANT: Blumenfeld, Marla

APPLICANT: Blumenfeld, Marla

APPLICANT: Blumenfeld, Marla

APPLICANT: Blumenfeld, Marla

APPLICANT: Blumenfeld, Marla

APPLICANT: Blumenfeld, Marla

APPLICANT: Blumenfeld, Marla

APPLICANT: Blumenfeld, Marla

APPLICANT: Blumenfeld, Marla

APPLICANT: Blumenfeld, Marla

APPLICANT: Blumenfeld, Marla

APPLICANT: Blumenfeld, Marla

APPLICANT: Blumenfeld, Marla

APPLICANT: Blumenfeld, Marla

APPLICANT: Blumenfeld, Marla

APPLICANT: Blumenfeld, Marla

APPLICANT: Blumenfeld, Marla

APPLICANT: Blumenfeld, Marla

APPLICANT: Blumenfeld, Marla

APPLICANT: Blumenfeld, Marla

APPLICANT: Blumenfeld, Marla

us-09-966-880a-10\_copy\_1\_300.ini

Qy 6 CTTGTAATCCAGTACTACAGAGGCTGAGGAGGATCCGGAGGAGCTGAGATCT 65

Db 217 CTTGTAATCCAGTACTACAGAGGCTGAGGAGGATCCGGAGGAGCTGAGATCT 177

Qy 66 GCTTGAAGCTGAGGAGGCTGAGGAGGATCCGGAGGAGCTGAGATCT 125

Db 176 GTTGAAGCTGAGGAGGATCCGGAGGAGCTGAGGAGGATCCGGAGGAGCTGAGATCT 117

Qy 126 GGGCGCAAGTGAAGCTTACCAAAAAA 161

Db 116 GGGCGCAAGTGAAGCTTACCAAAAAA 35289

## RESULT 7

US-08-781-891-79

Sequence 79, Application US/08781891

Patent No. 6090620

GENERAL INFORMATION:

APPLICANT: Pu, Ying-Hui

APPLICANT: Pu, Ying-Hui

APPLICANT: Pu, Ying-Hui

APPLICANT: Pu, Ying-Hui

APPLICANT: Pu, Ying-Hui

APPLICANT: Pu, Ying-Hui

APPLICANT: Pu, Ying-Hui

APPLICANT: Pu, Ying-Hui

APPLICANT: Pu, Ying-Hui

APPLICANT: Pu, Ying-Hui

APPLICANT: Pu, Ying-Hui

APPLICANT: Pu, Ying-Hui

APPLICANT: Pu, Ying-Hui

APPLICANT: Pu, Ying-Hui

APPLICANT: Pu, Ying-Hui

APPLICANT: Pu, Ying-Hui

APPLICANT: Pu, Ying-Hui

APPLICANT: Pu, Ying-Hui

APPLICANT: Pu, Ying-Hui

APPLICANT: Pu, Ying-Hui

APPLICANT: Pu, Ying-Hui

APPLICANT: Pu, Ying-Hui

APPLICANT: Pu, Ying-Hui

APPLICANT: Pu, Ying-Hui

APPLICANT: Pu, Ying-Hui

APPLICANT: Pu, Ying-Hui

APPLICANT: Pu, Ying-Hui

APPLICANT: Pu, Ying-Hui

APPLICANT: Pu, Ying-Hui

APPLICANT: Pu, Ying-Hui

APPLICANT: Pu, Ying-Hui

APPLICANT: Pu, Ying-Hui

APPLICANT: Pu, Ying-Hui

APPLICANT: Pu, Ying-Hui

APPLICANT: Pu, Ying-Hui

APPLICANT: Pu, Ying-Hui

APPLICANT: Pu, Ying-Hui

APPLICANT: Pu, Ying-Hui

APPLICANT: Pu, Ying-Hui

APPLICANT: Pu, Ying-Hui

APPLICANT: Pu, Ying-Hui

APPLICANT: Pu, Ying-Hui



QY 181 ATCAAGATC 189  
DB 42296 AGTTCATC 42304

## RESULT 8

US-09-618-166-79

; Sequence 79, Application US/09618166

; Patent No. 6583112

; GENERAL INFORMATION:

; APPLICANT: Fu, Ying-Hui

; Yu, Chang-En

; Oshima, Junko

; Mulligan, John T.

; Schellenberg, Gerald D.

; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO

; WERNER'S SYNDROME

; NUMBER OF SEQUENCES: 209

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Seed Intellectual Property Law Group

; STREET: 701 Fifth Avenue, Suite 6300

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/618,166

; FILING DATE: 17-Jul-2000

; CLASSIFICATION: &lt;Unknown&gt;

; ATTORNEY/AGENT INFORMATION:

; NAME: McMaesters, David D.

; REGISTRATION NUMBER: 33,963

; REFERENCE/DOCKET NUMBER: 240052.419C1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; INFORMATION FOR SEQ ID NO: 79:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 87350 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; SEQUENCE DESCRIPTION: SEQ ID NO: 79:

US-09-618-166-79

Query Match 26.9%; Score 80.8; DB 4; Length 87350;

Best Local Similarity 69.8%; Pred. No. 8.2e-12; Indels 19; Gaps 1;

Matches 132; Conservative 0; Mismatches 38; ----- 42179

QY 1 GGGGCGCTGTATCCAGCTACTACAGAGGCTGAGGAGGATCCGCGAGCTGGCA 60  
DB 42135 GGGGCGCTGTATCCAGCTACTACAGAGGCTGAGGAGGAT----- 42179

QY 61 GATTCGCTGAGCTGGAGGTTGAGGTTACGTAAGCCAGATCATGCCAGTACTTC 120  
DB 42180 -----TGCTTGAACCCAGAGAGGTGAGGTGAGTGAAGATTGCGCACTGCACTTC 42235

QY 121 AGCTGGGCGCAAGTGAGCCCTTAACAAAAAATTTAAATAAGAAATTAG 180  
DB 42236 AGCTGGGCGCAAGTGAGCCCTTCTGTCNAAAAAATTTAAATAAGAAATAA 42295

QY 181 ATCAAGATC 189  
DB 42296 AGTTCATC 42304

## RESULT 9

US-09-791-211-3

; Sequence 3, Application US/09791211

; Patent No. 6448080

; GENERAL INFORMATION:

; APPLICANT: Donna T. Ward

; Applicant: Andrew T. Watt

; TITLE OF INVENTION: ANTISENSE MODULATION OF WRN EXPRESSION

; FILE REFERENCE: RTS-0205

; CURRENT APPLICATION NUMBER: US/09/791,211

; NUMBER OF SEQ ID NOS: 90

; SEQ ID NO 3

; LENGTH: 87543

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: unsure

; LOCATION: 7421

; OTHER INFORMATION: unknown

; NAME/KEY: unsure

; LOCATION: 7427

; OTHER INFORMATION: unknown

; NAME/KEY: unsure

; LOCATION: 11609

; OTHER INFORMATION: unknown

; NAME/KEY: unsure

; LOCATION: 12605

; OTHER INFORMATION: unknown

; NAME/KEY: unsure

; LOCATION: 12742

; OTHER INFORMATION: unknown

; NAME/KEY: unsure

; LOCATION: 29370

; OTHER INFORMATION: unknown

; NAME/KEY: unsure

; LOCATION: 29422

; OTHER INFORMATION: unknown

; NAME/KEY: unsure

; LOCATION: 29979

; OTHER INFORMATION: unknown

; NAME/KEY: unsure

; LOCATION: 29980

; OTHER INFORMATION: unknown

; NAME/KEY: unsure

; LOCATION: 30135

; OTHER INFORMATION: unknown

; NAME/KEY: unsure

; LOCATION: 30140

; OTHER INFORMATION: unknown

; NAME/KEY: unsure

; LOCATION: 31205

; OTHER INFORMATION: unknown

; NAME/KEY: unsure

; LOCATION: 31206

; OTHER INFORMATION: unknown

; NAME/KEY: unsure

; LOCATION: 31095

; OTHER INFORMATION: unknown

; NAME/KEY: unsure

; LOCATION: 31592

; OTHER INFORMATION: unknown

; NAME/KEY: unsure

; LOCATION: 33160

; OTHER INFORMATION: unknown

; NAME/KEY: unsure

; LOCATION: 34066

; OTHER INFORMATION: unknown

; NAME/KEY: unsure

; LOCATION: 34072

; OTHER INFORMATION: unknown



NAME/KEY: unsure  
LOCATION: 36816  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 39020  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 42164  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 42459  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 46808  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 46823  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 46826  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 47291  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 52785  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 52787  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 53384  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 54684  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 59215  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 59235  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 59242  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 63290  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 66614  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 68660  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 68697  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 68718  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 68733  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 68739  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 69785  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 79134  
OTHER INFORMATION: unknown  
NAME/KEY: unsure

LOCATION: 79198  
OTHER INFORMATION: unknown  
NAME/KEY: unsure  
LOCATION: 86336  
OTHER INFORMATION: unknown  
US-09-791-211-3

Query Match 26.9%; Score 80.8; DB 4; Length 87543;  
Best Local Similarity 69.8%; Pred. No. 8.2e-12; Matches 132; Conservative 0; Mismatches 38; Indels 19; Gaps 1;

QY 1 GGGGCGCTGTAATCCAGCTACTCAGAGGCTGAGGCGAGAGATCCGGAGCCTGGCA 60  
DB 42328 GGGTCCCTGTAATCCAGCTACTCAGAGGCTGAGGCGAGAGAT----- 42372  
QY 61 GATTCGCTGAGCTGGAGGTTGAGGCTACAGTAAGCCAGATCATGCCATTACTTC 120  
DB 42373 ---TGCTTGAACCCGAGGAGTGAGGTTGCAGTGAATGTCCTGCACTGCACCTTC 42428  
QY 121 AGCTGGGCGACAAAGTGAGCCGTAACAAATAAATAAATAAATAAATAAATAA 180  
DB 42429 AGCTGGGAGACGAGCGAGCTCTGTCNNAAAAAAAAAAAAAAAAAGAGTAAAA 42488  
QY 181 ATCAAGATC 189  
DB 42489 AGGTACATC 42497

RESULT 10  
US-09-539-333D-1  
Sequence 1, Application US/0953933D  
Patent No. 6476208  
GENERAL INFORMATION:  
APPLICANT: Cohen, Daniel  
APPLICANT: Blumenfeld, Marta  
APPLICANT: Chumakov, Ilya  
APPLICANT: Bougueterec, Lydie  
APPLICANT: Bihaun, Bernard  
APPLICANT: Besiloux, Laurent  
TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLUELIC MARKERS  
FILE REFERENCE: GENSET.0474US  
CURRENT APPLICATION NUMBER: US/09/539,333D  
CURRENT FILING DATE: 2000-03-30  
PRIOR APPLICATION NUMBER: US 60/126,903  
PRIOR FILING DATE: 1999-03-30  
PRIOR APPLICATION NUMBER: US 60/131,971  
PRIOR FILING DATE: 1999-04-30  
PRIOR APPLICATION NUMBER: US 60/132,065  
PRIOR FILING DATE: 1999-04-30  
PRIOR APPLICATION NUMBER: US 60/143,928  
PRIOR FILING DATE: 1999-07-14  
PRIOR APPLICATION NUMBER: US 60/145,915  
PRIOR FILING DATE: 1999-07-27  
PRIOR APPLICATION NUMBER: US 60/146,453  
PRIOR FILING DATE: 1999-07-29  
PRIOR APPLICATION NUMBER: US 60/146,452  
PRIOR FILING DATE: 1999-07-29  
PRIOR APPLICATION NUMBER: US 60/162,288  
PRIOR FILING DATE: 1999-10-28  
PRIOR APPLICATION NUMBER: US 09/416,384  
PRIOR FILING DATE: 1999-10-12  
NUMBER OF SEQ ID NOS: 231  
SOFTWARE: Patent.pm  
SEQ ID NO 1  
LENGTH: 319608  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 31..1107  
OTHER INFORMATION: 5'regulatory region g35018 gene  
FEATURE:



NAME/KEY: exon  
LOCATION: 1108..1289  
OTHER INFORMATION: exon A g35018 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 14877..14920  
OTHER INFORMATION: exon B g35018 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 18778..18862  
OTHER INFORMATION: exon Bbis g35018 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 25593..25740  
OTHER INFORMATION: exon C g35018 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 29388..29502  
OTHER INFORMATION: exon D g35018 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 29967..30282  
OTHER INFORMATION: exon E g35018 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 65505..65853  
OTHER INFORMATION: exon G g35018 gene  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 65854..67854  
OTHER INFORMATION: 3' regulatory region g35018 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 94124..94964  
OTHER INFORMATION: exon g35017  
FEATURE:  
NAME/KEY: exon  
LOCATION: 201188..201234  
OTHER INFORMATION: exon S g35030 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 214676..214793  
OTHER INFORMATION: exon T g35030 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 215702..215746  
OTHER INFORMATION: exon U g35030 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 216836..216915  
OTHER INFORMATION: exon V g35030 gene  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 213818..215818  
OTHER INFORMATION: 3' regulatory region g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 215819..215941  
OTHER INFORMATION: exon R complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 215819..215975  
OTHER INFORMATION: exon Rbis complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 216661..216952  
OTHER INFORMATION: exon Qbis complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 216661..217061  
OTHER INFORMATION: exon Q complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 217027..217061  
OTHER INFORMATION: exon Q1 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 229647..229742  
OTHER INFORMATION: exon X complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 230408..230721  
OTHER INFORMATION: exon P complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 231272..231412  
OTHER INFORMATION: exon Obis complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 231787..231880  
OTHER INFORMATION: exon O2 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 231870..231879  
OTHER INFORMATION: exon O1 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 234174..234321  
OTHER INFORMATION: exon O complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 237406..237428  
OTHER INFORMATION: exon Nbis complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 239719..239807  
OTHER INFORMATION: exon N2 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 239719..239853  
OTHER INFORMATION: exon N complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 240528..240569  
OTHER INFORMATION: exon M1117 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 240528..240596  
OTHER INFORMATION: exon M1090 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 240528..240617  
OTHER INFORMATION: exon M1069 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 240528..240644  
OTHER INFORMATION: exon M2 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 240528..240924  
OTHER INFORMATION: exon M692 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 240528..241685  
OTHER INFORMATION: exon M1 complement g34872 gene  
FEATURE:  
NAME/KEY: exon  
LOCATION: 240600..240993



```

; OTHER INFORMATION: exon MS1 complement g34872 gene
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 241686..243685
; OTHER INFORMATION: 5' regulatory region g34872 gene
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 290652..292652
; OTHER INFORMATION: 3' regulatory region g34665 gene
; FEATURE:
; NAME/KEY: exon
; LOCATION: 292653..292841

Query Match      26.9%; Score 80.8; DB 4; Length 319608;
Best Local Similarity 63.6%; Pred. No. 1.3e-11;
Matches 150; Conservative 0; Mismatches 67; Indels 19; Gaps 1;

QY      2 GGGGCGCTGTATCCGAGCTACTCAGAGAGCTGAGGAGAGGATCCGGAGCCTGGCAG 61
        |||
DB      191544 GGCACCTGTAAATACCACTACTCCGGAGGCTGAGGACAGATCA----- 191589
        |||

QY      62 ATCTGCGCTGAGCGCTGGAGTTGAGGCTACAGTACCAAGATCATGCGCAGTATACCTCA 121
        |||
DB      191590 -----CTTGACCTGGAGATGAGAGTTGCAGTGCAGATCATGCGCAGTACTTCA 191644
        |||

QY      122 GCGCTGGCGCAAAAGTGGAGCGGTAACAAAAAAAAAATTTAAAAAGAAATTAGA 181
        |||
DB      191645 GCGTGGACAAACAAAGCAAGCTCTGTCTCAAAAAATTTAAAAATTTAAAAA 191704
        |||

QY      182 TCAAGATCCAACTGTAAAAAGTGGCTTAACACCAATTAAGAGTTGAGTTA 237
        |||
DB      191705 AATAAAGGAATCCACAGATTGTATTCTTTAATACTAGATTAGTGTCTA 191760
        |||

RESULT 11
US-09-679-409-1
; Sequence 1, Application US/09679409
; Patent No. 6553316
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Blumentfeld, Marta
; APPLICANT: Chumakov, Ilya
; APPLICANT: Bouguerelet, Lydie
; APPLICANT: Besieux, Laurent
; TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENE, PROTEINS AND BIALLELIC MARKERS
; FILE REFERENCE: 53.US15.CIP
; CURRENT FILING DATE: US/09/679,409
; PRIOR APPLICATION NUMBER: 09/539,333
; PRIOR FILING DATE: 2000-03-03
; PRIOR APPLICATION NUMBER: 09/416,384
; PRIOR FILING DATE: 1999-10-12
; PRIOR APPLICATION NUMBER: 60/168,088
; PRIOR FILING DATE: 1999-11-30
; NUMBER OF SEQ ID NOS: 134
; SOFTWARE: Patent.prm
; SEQ ID NO 1
; LENGTH: 319608
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 199122..201122
; OTHER INFORMATION: 5' regulatory region
; NAME/KEY: exon
; LOCATION: 201123..201234
; OTHER INFORMATION: exon S
; NAME/KEY: exon
; LOCATION: 201123..201560
; OTHER INFORMATION: exon S2
; NAME/KEY: exon
; LOCATION: 214676..214793
; OTHER INFORMATION: exon T

; NAME/KEY: exon
; LOCATION: 215702..215746
; OTHER INFORMATION: exon U
; NAME/KEY: exon
; LOCATION: 216836..216984
; OTHER INFORMATION: exon V
; NAME/KEY: exon
; LOCATION: 216836..217077
; OTHER INFORMATION: exon V2
; NAME/KEY: exon
; LOCATION: 217671..217764
; OTHER INFORMATION: exon V1
; NAME/KEY: exon
; LOCATION: 227655..227736
; OTHER INFORMATION: exon V4
; NAME/KEY: exon
; LOCATION: 238715..238919
; OTHER INFORMATION: exon V3
; NAME/KEY: exon
; LOCATION: 240440..240673
; OTHER INFORMATION: exon W
; NAME/KEY: exon
; LOCATION: 241072..241291
; OTHER INFORMATION: exon W2
; NAME/KEY: exon
; LOCATION: 244353..244561
; OTHER INFORMATION: exon X
; NAME/KEY: exon
; LOCATION: 246273..247802
; OTHER INFORMATION: exon Y
; NAME/KEY: exon
; LOCATION: 247803..249803
; OTHER INFORMATION: exon Z
; NAME/KEY: misc.feature
; LOCATION: 247803..249803
; OTHER INFORMATION: 3' regulatory region
; NAME/KEY: allele
; LOCATION: 8316
; OTHER INFORMATION: 99-27943-150 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 21672
; OTHER INFORMATION: 99-27935-193 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 65485
; OTHER INFORMATION: 8-128-33 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 95396
; OTHER INFORMATION: 99-31960-363 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 107281
; OTHER INFORMATION: 99-24656-260 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 160640
; OTHER INFORMATION: 99-24639-163 : polymorphic base A or C
; NAME/KEY: allele
; LOCATION: 160876
; OTHER INFORMATION: 99-24634-108 : polymorphic base A or T
; NAME/KEY: allele
; LOCATION: 168974
; OTHER INFORMATION: 99-7652-162 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 170810
; OTHER INFORMATION: 99-16100-147 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 173358
; OTHER INFORMATION: 99-5862-167 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 189957
; OTHER INFORMATION: 99-5919-215 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 197163
; OTHER INFORMATION: 99-24658-410 : polymorphic base A or G
; NAME/KEY: allele
```



LOCATION: 200778 : polymorphic base A or G  
OTHER INFORMATION: 8-303-235 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 202651 : polymorphic base A or G  
OTHER INFORMATION: 8-300-221 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 202679 : polymorphic base A or G  
OTHER INFORMATION: 8-300-193 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 203378 : polymorphic base A or T  
OTHER INFORMATION: 8-299-128 : polymorphic base A or T  
NAME/KEY: allele  
LOCATION: 204138 : polymorphic base A or T  
OTHER INFORMATION: 8-296-213 : polymorphic base A or T  
NAME/KEY: allele  
LOCATION: 204605 : polymorphic base C or T  
OTHER INFORMATION: 8-252-190 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 204934 : polymorphic base A or G  
OTHER INFORMATION: 99-24644-194 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 205206 : polymorphic base A or C  
OTHER INFORMATION: 8-295-248 : polymorphic base A or C  
NAME/KEY: allele  
LOCATION: 205329 : polymorphic base C or T  
OTHER INFORMATION: 8-295-125 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 206064 : polymorphic base A or G  
OTHER INFORMATION: 8-293-130 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 206545 : polymorphic base A or G  
OTHER INFORMATION: 8-292-198 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 207313 : polymorphic base A or G  
OTHER INFORMATION: 8-251-322 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 208285 : polymorphic base A or G  
OTHER INFORMATION: 8-289-322 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 208960 : polymorphic base C or T  
OTHER INFORMATION: 8-287-249 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 209123 : polymorphic base A or T  
OTHER INFORMATION: 8-287-86 : polymorphic base A or T  
NAME/KEY: allele  
LOCATION: 209631 : polymorphic base A or G  
OTHER INFORMATION: 8-285-319 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 210361 : polymorphic base G or C  
OTHER INFORMATION: 8-283-278 : polymorphic base G or C  
NAME/KEY: allele  
LOCATION: 210463 : polymorphic base A or G  
OTHER INFORMATION: 8-283-176 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 210486 : polymorphic base G or C  
OTHER INFORMATION: 8-283-153 : polymorphic base G or C  
NAME/KEY: allele  
LOCATION: 210583 : polymorphic base C or T  
OTHER INFORMATION: 8-283-56 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 210879 : polymorphic base G or C  
OTHER INFORMATION: 8-282-345 : polymorphic base G or C  
NAME/KEY: allele  
LOCATION: 210964 : polymorphic base G or T  
OTHER INFORMATION: 8-282-260 : polymorphic base G or T  
NAME/KEY: allele  
LOCATION: 210979 : polymorphic base A or C  
OTHER INFORMATION: 8-282-245 : polymorphic base A or C  
NAME/KEY: allele  
LOCATION: 211050 : polymorphic base A or C  
OTHER INFORMATION: 8-282-174 : variable motif AAAG or GAAGGAAGGAAGGAAGA  
NAME/KEY: allele  
LOCATION: 211132

OTHER INFORMATION: 8-282-92 : polymorphic base A or T  
NAME/KEY: allele  
LOCATION: 211247 : polymorphic base A or G  
OTHER INFORMATION: 8-281-367 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 211315 : polymorphic base A or G  
OTHER INFORMATION: 8-281-299 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 211366 : polymorphic base G or C  
OTHER INFORMATION: 8-281-248 : polymorphic base G or C  
NAME/KEY: allele  
LOCATION: 212520 : polymorphic base A or C  
OTHER INFORMATION: 8-279-197 : polymorphic base A or C  
NAME/KEY: allele  
LOCATION: 212821 : polymorphic base C or T  
OTHER INFORMATION: 8-278-289 : polymorphic base C or T  
NAME/KEY: allele

Query Match 26.9%; Score 80.6; DB 4; Length 319608;  
Best Local Similarity 63.6%; Pred. No. 1.3e-11;  
Matches 150; Conservative 0; Mismatches 67; Indels 19; Gaps 1;

QY 2 GGGGCGCTGTAATCCAGCTACTCAGAGCTGAGGACAGAGATCCGGAGCTGCGAG 61  
DB 191544 GGACCTGTATATACAGCTACTCGGAGCTGAGGACAGAGATCA----- 191589  
QY 62 ATCTGCTGAGCGCTGAGGTTGAGCTACAGTAAGCCAGATATGCGCAGTATCTTCA 121  
DB 191590 -----CTTGAACCTGGAGATGAGGTTGAGGAGTATGCGAGATATGCGCAGTATCTTCA 191644  
QY 122 GCGTGGGACAAAGTGAAGCCGTAAACAAAAAATTTAAATAAAGAAATTTGA 181  
DB 191645 GCGTGGACAAAGTGAAGCCGTAAACAAAAAATTTAAATAAATAAATAAATAA 191704  
QY 182 TCAGATCCAACTGTAATTAAGTGGCTTAACCCACATTAAGGTTTGAAGTTA 237  
DB 191705 AATTAAGTGAATCCAGAGATTTGCTTATTTATTTATTAAGTATGATGCTCA 191760

## RESULT 12

US-09-056-105-31  
Sequence 31, Application US/09056105  
Patent No. 6287569  
GENERAL INFORMATION:  
APPLICANT: KIPPS, THOMAS J.  
APPLICANT: KU, YUNOJ  
TITLE OF INVENTION: VACCINES WITH ENHANCED INTRACELLULAR  
FILE REFERENCE: 233/221  
CURRENT APPLICATION NUMBER: US/09/056,105  
CURRENT FILING DATE: 1998-04-06  
EARLIER APPLICATION NUMBER: 60/043,467  
EARLIER FILING DATE: 1997-04-10  
NUMBER OF SEQ ID NOS: 35  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 31  
LENGTH: 7130  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-056-105-31

Query Match 26.9%; Score 80.6; DB 3; Length 7130;  
Best Local Similarity 66.0%; Pred. No. 3.9e-12;  
Matches 142; Conservative 0; Mismatches 54; Indels 19; Gaps 1;

QY 5 GCGTGTATCCAGCTACTCAGAGCTGAGGACAGAGATCCGGAGCTGCGAGATC 64  
DB 2755 GCGTGTATCCAGCTACTCAGAGCTGAGGACAGAGAT----- 2795  
QY 65 TCGCTAGCGCTGGAGGTTGAGGCTACAGTAAGCCAAAGTCAATGCGCAGTATCTTCA 124  
DB 2796 TCGCTAGCGCTGGAGGTTGAGGCTACAGTAAGCCAAAGTCAATGCGCAGTATCTTCA 2855







GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 12, 2004, 21:22:51 ; Search time 105.276 Seconds

(without alignments)  
10491.276 Million cell updates/sec

Title: US-09-966-880A-10\_COPY\_1\_300

Perfect score: 300

Sequence: 1 999999ccgttaaccacgct.....aatgcatgtgcacctcgt 300

Scoring table: IDENTITY NUC

Searched: 2432557 seqs, 1840798884 residues

Total number of hits satisfying chosen parameters: 4865114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09C\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10C\_NEW\_PUB.seq:\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10C\_NEW\_PUB.seq:\*
- 17: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 18: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	300	100.0	6564	9	US-09-966-880A-10
2	298.4	99.5	11204	9	US-09-966-880A-35
3	100	33.3	3523	15	US-10-027-632-114084
4	94	31.3	32134	15	US-09-764-991-6303
5	94	31.3	32134	14	US-10-205-428-608
6	94	31.3	32191	10	US-09-764-891-6304
7	94	31.3	32191	14	US-10-205-428-609
8	93.4	31.1	669	15	US-10-027-632-284884
9	93.2	31.1	143899	10	US-09-972-546-15
10	91.4	30.5	1859	11	US-09-984-429-549
11	91.4	30.5	12437	14	US-10-314-321A-3
12	91	30.3	849	15	US-10-027-632-29551
13	91	30.3	849	15	US-10-027-632-29552
14	91	30.3	56737	9	US-09-782-378A-17
15	90.6	30.2	595	15	US-10-027-632-223820

C 16	89.8	29.9	13893	13	US-10-161-510-1	Sequence 1, Appli
C 17	89	29.7	1236	15	US-10-027-632-124822	Sequence 124822,
C 18	89	29.7	1252	15	US-10-027-632-257891	Sequence 257891,
C 19	89	29.7	1252	15	US-10-027-632-257892	Sequence 257892,
C 20	89	29.7	1252	15	US-10-027-632-257893	Sequence 257893,
C 21	89	29.7	1252	15	US-10-027-632-257894	Sequence 257894,
C 22	89	29.7	1252	15	US-10-027-632-257895	Sequence 257895,
C 23	88.4	29.5	483	15	US-10-027-632-275643	Sequence 275643,
C 24	88	29.3	485	15	US-10-027-632-21405	Sequence 21405, A
C 25	88	29.3	485	15	US-10-027-632-21406	Sequence 21406, A
C 26	88	29.3	846	15	US-10-027-632-160172	Sequence 160172,
C 27	88	29.3	846	15	US-10-027-632-160173	Sequence 160173,
C 28	88	29.3	846	15	US-10-027-632-160174	Sequence 160174,
C 29	87.8	29.3	3157	15	US-10-027-632-115461	Sequence 115461,
C 30	87.8	29.3	6318	15	US-10-247-335-1852	Sequence 882, App
C 31	87.8	29.3	13069	9	US-09-764-869-1850	Sequence 1850, App
C 32	87.8	29.3	13069	14	US-10-091-504-1850	Sequence 1850, App
C 33	87.8	29.3	13069	15	US-10-227-577-1850	Sequence 1849, App
C 34	87.8	29.3	32203	9	US-09-764-869-1849	Sequence 1849, App
C 35	87.8	29.3	32203	14	US-10-091-504-1849	Sequence 1849, App
C 36	87.8	29.3	32203	15	US-10-227-577-1849	Sequence 1849, App
C 37	87.6	29.2	2445	15	US-10-027-632-110185	Sequence 110185,
C 38	87.4	29.1	668	15	US-10-027-632-131147	Sequence 131147,
C 39	86.6	28.9	473	15	US-10-027-632-106241	Sequence 106241,
C 40	86.4	28.8	2716	15	US-10-027-632-111585	Sequence 111585,
C 41	86.4	28.8	2716	15	US-10-027-632-111586	Sequence 111586,
C 42	86.2	28.7	482	15	US-10-027-632-70410	Sequence 70410, A
C 43	86.2	28.7	482	15	US-10-027-632-108468	Sequence 108468,
C 44	86.2	28.7	482	15	US-10-027-632-294298	Sequence 294298,
C 45	85.8	28.6	497	10	US-09-918-995-21543	Sequence 21543, A

## ALIGNMENTS

US-09-966-880A-10	RESULT 1
Sequence 10, Application US/09966880A	
Patent No. US20020164743A1	
GENERAL INFORMATION:	
APPLICANT: Honjo, Tasuku	
APPLICANT: Muramatsu, Masamichi	
TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE	
FILE REFERENCE: 06501-088001	
CURRENT APPLICATION NUMBER: US/09/966, 880A	
PRIOR FILING DATE: 2001-09-28	
PRIOR APPLICATION NUMBER: PCT/JP00/01918	
PRIOR FILING DATE: 2000-03-28	
PRIOR APPLICATION NUMBER: JP 11-371382	
PRIOR FILING DATE: 1999-12-27	
PRIOR APPLICATION NUMBER: JP 11-178999	
PRIOR FILING DATE: 1999-06-24	
PRIOR APPLICATION NUMBER: JP 11-87192	
PRIOR FILING DATE: 1999-03-29	
NUMBER OF SEQ ID NOS: 36	
SOFTWARE: FASTSEQ for Windows Version 4.0	
SEQ ID NO 10	
LENGTH: 6564	
TYPE: DNA	
ORGANISM: Homo sapiens	
US-09-966-880A-10	
Query Match	100.0%; Score 300; DB 9; Length 6564;
Best Local Similarity	100.0%; Pred. No. 1.2e-65;
Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
DB	1 GGGGGCCTGATATCCAGCTACTACGAGAGCTGAGGACGAGAGATCCGGAGCCTGGGA 60
QY	1 GGGGGCCTGATATCCAGCTACTACGAGAGCTGAGGACGAGAGATCCGGAGCCTGGGA 60
DB	1 GGGGGCCTGATATCCAGCTACTACGAGAGCTGAGGACGAGAGATCCGGAGCCTGGGA 60
QY	1 GATTCGCTGAGGCTGGGAGGTTGAGGCTACGATAGCAAGATCATGCTCATATCTTC 120
DB	1 GATTCGCTGAGGCTGGGAGGTTGAGGCTACGATAGCAAGATCATGCTCATATCTTC 120



QY 121 AGCCTGGGCGCAAGTGAAGCCTTAACAAAAAATTTAAAAAAGAATTAG 180  
DB 121 AGCCTGGGCGCAAGTGAAGCCTTAACAAAAAATTTAAAAAAGAATTAG 180  
QY 181 ATCAAGATCCAACTGTATTAAGAGTGGCTTAACACCACTTAAGAGTTTATTC 240  
DB 181 ATCAAGATCCAACTGTATTAAGAGTGGCTTAACACCACTTAAGAGTTTATTC 240  
QY 241 TGCAGGCGAAGAGAACATCAAGGAGGCTTTCAGCATTTGGCATTTGGCACTTGT 300  
DB 241 TGCAGGCGAAGAGAACATCAAGGAGGCTTTCAGCATTTGGCATTTGGCACTTGT 300

## RESULT 2

US-09-966-880A-35  
Sequence 35, Application US/09966880A  
Patent No. US2002016473A1  
GENERAL INFORMATION:  
APPLICANT: Honjo, Tasuku  
APPLICANT: Matsumatsu, Masamichi  
TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
FILE REFERENCE: 06501-088001  
CURRENT APPLICATION NUMBER: US/09/966,880A  
CURRENT FILING DATE: 2001-09-28  
PRIOR APPLICATION NUMBER: PCT/JP00/01918  
PRIOR FILING DATE: 2000-03-28  
PRIOR APPLICATION NUMBER: JP 11-371382  
PRIOR FILING DATE: 1999-12-27  
PRIOR APPLICATION NUMBER: JP 11-178999  
PRIOR FILING DATE: 1999-06-24  
PRIOR APPLICATION NUMBER: JP 11-87192  
PRIOR FILING DATE: 1999-03-29  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 35  
LENGTH: 11204  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-966-880A-35

Query Match 99.5%; Score 298.4; DB 9; Length 11204;  
Best Local Similarity 99.7%; Pred. No. 3,6e-65;  
Matches 299; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGGGCGCTGTATTCCTAGCTACTCAGAGGCTGAGGAGGAGATCCGCGAGCCTTGCA 60  
DB 5216 GGGGCGCTGTATTCCTAGCTACTCAGAGGCTGAGGAGGAGATCCGCGAGCCTTGCA 5275  
QY 61 GATCTGCTGAGCCTGGAGGTTGAGGCTACAGTAAGCCAGATATCCAGTATCTTC 120  
DB 5276 GATCTGCTGAGCCTGGAGGTTGAGGCTACAGTAAGCCAGATATCCAGTATCTTC 5335  
QY 121 AACCTGGGGAAGAGTGAAGCCGTAAACAAAAAATTTAAAAAAGAATTAG 180  
DB 5336 AACCTGGGGAAGAGTGAAGCCGTAAACAAAAAATTTAAAAAAGAATTAG 5395  
QY 181 ATCAAGATCCAACTGTATTAAGAGTGGCTTAACACCACTTAAGAGTTTATTC 240  
DB 5396 ATCAAGATCCAACTGTATTAAGAGTGGCTTAACACCACTTAAGAGTTTATTC 5455  
QY 241 TGCAGGCGAAGAGAACATCAAGGAGGCTTTCAGCATTTGGCATTTGGCACTTGT 300  
DB 5456 TGCAGGCGAAGAGAACATCAAGGAGGCTTTCAGCATTTGGCATTTGGCACTTGT 5515

RESULT 3  
US-10-027-632-114084  
Sequence 114084, Application US/10027632  
Publication No. US20030204075A9  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide

FILE OF INVENTION: Polymorphisms in the Human Genome  
FILE REFERENCE: 108827,129  
CURRENT APPLICATION NUMBER: US/10/027,632  
CURRENT FILING DATE: 2002-04-30  
PRIOR APPLICATION NUMBER: US 60/218,006  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/198,676  
PRIOR FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: US 60/193,483  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: US 60/185,218  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/167,363  
PRIOR FILING DATE: 1999-11-23  
PRIOR APPLICATION NUMBER: US 60/156,358  
PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: US 60/146,002  
PRIOR FILING DATE: 1999-08-09  
NUMBER OF SEQ ID NOS: 325720  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 114084  
LENGTH: 3523  
TYPE: DNA  
ORGANISM: Human  
US-10-027-632-114084

Query Match 33.3%; Score 100; DB 15; Length 3523;  
Best Local Similarity 73.8%; Pred. No. 2.9e-15;  
Matches 127; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 3 GGGGCTGTATTCCTAGCTACTCAGAGGCTGAGGAGGAGATCCGCGAGCCTTGCA 62  
DB 2365 GGGGCTGTATTCCTAGCTACTCAGAGGCTGAGGAGGAGATCCGCGAGCCTTGCA 2424  
QY 63 TGTGCTGAGCCTGGAGGTTGAGGCTACAGTAAGCCAGATATCCAGTATCTTCAG 122  
DB 2425 TGTGCTGAGCCTGGAGGTTGAGGCTACAGTAAGCCAGATATCCAGTATCTTCAG 2484  
QY 123 CCTGGGCGAAGAGTGAAGCCGTAAACAAAAAATTTAAAAAAGA 174  
DB 2485 CCTGGGCGAAGAGTGAAGCCGTAAACAAAAAATTTAAAAAAGA 2536

RESULT 4  
US-09-764-891-6303  
Sequence 6303, Application US/09764891  
Publication No. US20030077808A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: FC006  
CURRENT APPLICATION NUMBER: US/09/764,891  
CURRENT FILING DATE: 2001-01-17  
Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 10231  
SOFTWARE: Patent Ver. 2.0  
SEQ ID NO 6303  
LENGTH: 32134  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-764-891-6303

Query Match 31.3%; Score 94; DB 10; Length 32134;  
Best Local Similarity 71.3%; Pred. No. 2.2e-13;  
Matches 124; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 1 GGGGCGCTGTATTCCTAGCTACTCAGAGGCTGAGGAGGAGATCCGCGAGCCTTGCA 60  
DB 12455 GGGGCGCTGTATTCCTAGCTACTCAGAGGCTGAGGAGGAGATCCGCGAGCCTTGCA 12514  
QY 61 GATCTGCTGAGCCTGGAGGTTGAGGCTACAGTAAGCCAGATATCCAGTATCTTC 120  
DB 12515 GATCTGCTGAGCCTGGAGGTTGAGGCTACAGTAAGCCAGATATCCAGTATCTTC 12574



QY 121 AGCTGGGCGCAAAAGTGAGACCGCTAACAAAAAATTTAAAAAGAA 174  
DB 12575 AGCTGGGTGACAAAGTGAGACTCCGCTCAAAAAAATTTAAAAATAA 12628

## RESULT 5

US-10-205-428-608  
Sequence 608, Application US/10205428  
Publication No. US20030108907A1  
GENERAL INFORMATION:

APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: P1117C1  
CURRENT APPLICATION NUMBER: US/10/205,428  
CURRENT FILING DATE: 2002-07-26

PRIOR APPLICATION NUMBER: 09/764,892  
PRIOR FILING DATE: 2001-01-17  
PRIOR APPLICATION NUMBER: 60/179,065  
PRIOR FILING DATE: 2000-01-31  
PRIOR APPLICATION NUMBER: 60/180,628  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: 60/214,886  
PRIOR FILING DATE: 2000-06-28  
PRIOR APPLICATION NUMBER: 60/217,487  
PRIOR FILING DATE: 2000-07-11  
PRIOR APPLICATION NUMBER: 60/225,758  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/220,963  
PRIOR FILING DATE: 2000-07-26  
PRIOR APPLICATION NUMBER: 60/217,496  
PRIOR FILING DATE: 2000-07-11  
PRIOR APPLICATION NUMBER: 60/225,447  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/218,290  
PRIOR FILING DATE: 2000-07-14  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 1019  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 608  
LENGTH: 32134  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-205-428-608

Query Match 31.3%; Score 94; DB 14; Length 32134;  
Best Local Similarity 71.3%; Pred. No. 2.2e-13;  
Matches 124; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 1 GGGGGGCTGTATATCCGACTCTACAGAGGCTGAGGAGAGAGATCCGGAGCTGGCA 60  
DB 12455 GGGGACCTGTACTCCGACTCTCGGAGGCTGAGGAGAGATGCGATGAGGAGGA 12514  
QY 61 GATCTGCTGAGCTGGAGGTTGAGGCTACAGTAAGCCAAAGATCATGCGATATCTTC 120  
DB 12515 GAATGGATGAACCCAGAGGAGAGGCTTCAGTGAAGCCAGATACACCATTCACCTCC 12574  
QY 121 AGCTGGGCGCAAAAGTGAGACCGCTAACAAAAAATTTAAAAAGAA 174  
DB 12575 AGCTGGGTGACAAAGTGAGACTCCGCTCAAAAAAATTTAAAAATAA 12628

## RESULT 6

US-09-764-891-6304  
Sequence 6304, Application US/09764891  
Publication No. US20030077808A1  
GENERAL INFORMATION:

APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PC006  
CURRENT APPLICATION NUMBER: US/09/764,891  
CURRENT FILING DATE: 2001-01-17

Prior application data removed - consult PALM or file wrapper

NUMBER OF SEQ ID NOS: 10231  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 6304  
LENGTH: 32191  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-764-891-6304

Query Match 31.3%; Score 94; DB 10; Length 32191;  
Best Local Similarity 71.3%; Pred. No. 2.2e-13;  
Matches 124; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 1 GGGGGGCTGTATATCCGACTCTACAGAGGCTGAGGAGAGATCCGGAGCTGGCA 60  
DB 9503 GGGGACCTGTACTCCGACTCTCGGAGGCTGAGGAGAGATGCGATGAGGAGCA 9562  
QY 61 GATCTGCTGAGCTGGAGGTTGAGGCTACAGTAAGCCAAAGATCATGCGATATCTTC 120  
DB 9563 GAATGGATGAACCCAGAGGAGAGGCTTCAGTGAAGCCAGATCACCATTCGACTCC 9622  
QY 121 AGCTGGGCGCAAAAGTGAGACCGCTAACAAAAAATTTAAAAAGAA 174  
DB 9623 AGCTGGGTGACAAAGTGAGACTCCGCTCAAAAAAATTTAAAAATAA 9676

## RESULT 7

US-10-205-428-609  
Sequence 609, Application US/10205428  
Publication No. US20030108907A1  
GENERAL INFORMATION:

APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: P1117C1  
CURRENT APPLICATION NUMBER: US/10/205,428  
CURRENT FILING DATE: 2002-07-26  
PRIOR APPLICATION NUMBER: 09/764,892  
PRIOR FILING DATE: 2001-01-17  
PRIOR APPLICATION NUMBER: 60/179,065  
PRIOR FILING DATE: 2000-01-31  
PRIOR APPLICATION NUMBER: 60/180,628  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: 60/214,886  
PRIOR FILING DATE: 2000-06-28  
PRIOR APPLICATION NUMBER: 60/217,487  
PRIOR FILING DATE: 2000-07-11  
PRIOR APPLICATION NUMBER: 60/225,758  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/220,963  
PRIOR FILING DATE: 2000-07-26  
PRIOR APPLICATION NUMBER: 60/217,496  
PRIOR FILING DATE: 2000-07-11  
PRIOR APPLICATION NUMBER: 60/225,447  
PRIOR FILING DATE: 2000-08-14  
PRIOR APPLICATION NUMBER: 60/218,290  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 1019  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 609  
LENGTH: 32191  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-205-428-609

Query Match 31.3%; Score 94; DB 14; Length 32191;  
Best Local Similarity 71.3%; Pred. No. 2.2e-13;  
Matches 124; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 1 GGGGGGCTGTATATCCGACTCTACAGAGGCTGAGGAGAGATCCGGAGCTGGCA 60  
DB 9503 GGGGACCTGTACTCCGACTCTCGGAGGCTGAGGAGAGATGCGATGAGGAGCA 9562  
QY 61 GATCTGCTGAGCTGGAGGTTGAGGCTACAGTAAGCCAAAGATCATGCGATATCTTC 120







```
; Sequence 549, Application US/09984429
; Publication No. US20040010132A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 53 Human Secreted Proteins
; FILE REFERENCE: P2018P2
; CURRENT APPLICATION NUMBER: US/09/984,429
; PRIOR FILING DATE: 2001-10-30
; PRIOR APPLICATION NUMBER: 60/244,591
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 09/288,143
; PRIOR FILING DATE: 1999-04-08
; PRIOR APPLICATION NUMBER: PCT/US98/21142
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 60/061,463
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/061,529
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/071,498
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/061,527
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/061,536
; PRIOR FILING DATE: 1997-10-09
; PRIOR APPLICATION NUMBER: 60/061,532
; NUMBER OF SEQ ID NOS: 727
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 549
; LENGTH: 1859
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-984-429-549

Query Match
Best Local Similarity 30.5%; Score 91.4; DB 11; Length 1859;
Matches 119; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 3 GGGCTGTAATCCAGCTACTCAGAGGCTGAGGAGGATCCGGAGCCTGGCAGA 62
Db 165 GCGGCTGTAGCCAGCTATTCAGAGGCTGAGGAGGATCCGGAGCCTGGCAGA 106
QY 63 TCTGCTGAGGCTGAGGCTGAGGCTGAGGAGGATCCGGAGCCTGGCAGA 122
Db 105 ATTCTTGAACTCCGAGGAGGCTGAGGAGGATCCGGAGCCTGGCAGA 46
QY 123 CTTGGGCGACAAAGTGAAGCCGTAACAAAAAAAAAATTTAA 167
Db 45 TCTAGGCAACGCGCAAACTCCATCTCCAAAAAATTTAA 1

RESULT 11
US-10-314-321A-3
; Sequence 3, Application US/10314321A
; Publication No. US20030190648A1
; GENERAL INFORMATION:
; APPLICANT: Hitachi, Ltd.
; TITLE OF INVENTION: Gene Predicting Method
; FILE REFERENCE: 31010118US1
; CURRENT APPLICATION NUMBER: US/10/314,321A
; PRIOR FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: JP 2002-103333
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 12437
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-314-321A-3

Query Match
Best Local Similarity 30.5%; Score 91.4; DB 14; Length 12437;
Matches 72.1%; Pred. No. 7e-13;
```

```
Matches 119; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 5 GCGTGTAAATCCAGCTACTCAGAGGCTGAGGAGGATCCGGAGCCTGGCAGATC 64
Db 3122 GCGTGTAAATCCAGCTACTCAGAGGCTGAGGAGGATCCGGAGCCTGGCAGATC 3181
QY 65 TCCCTGAGCCTGGAGGCTTGAAGCTACAGTAAAGCCAAATATGCAATATCTTACGCC 124
Db 3182 CTTTGAAGCAGTGGGAGGCTTGAAGCTACAGTAAAGCCAAATATGCAATATCTTACGCC 3241
QY 125 TGGGGGCAAAAGTGAAGCCGTAACAAAAAAAAAATTTAA 169
Db 3242 TGGGTGACAGCAAGCAACCTGTCTCAAAAAAACAACCA 3286

RESULT 12
US-10-027-632-29551
; Sequence 29551, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29551
; LENGTH: 849
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-29551

Query Match
Best Local Similarity 30.3%; Score 91; DB 15; Length 849;
Matches 121; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 4 GCGTGTAAATCCAGCTACTCAGAGGCTGAGGAGGATCCGGAGCCTGGCAGAT 63
Db 410 GCGTGTAAATCCAGCTACTCAGAGGCTGAGGAGGATCCGGAGCCTGGCAGAT 469
QY 64 TCGCTGAGCCTGGAGGCTTGAAGCTACAGTAAAGCCAAATATGCAATATCTTACGCC 123
Db 470 TCGCTGAGCCTGGAGGCTTGAAGCTACAGTAAAGCCAAATATGCAATATCTTACGCC 529
QY 124 CTTGGGCGACAAAGTGAAGCCGTAACAAAAAAAAAATTTAA 174
Db 530 CTTGGGCGACAAAGTGAAGCCGTAACAAAAAAAAAATTTAA 580

RESULT 13
US-10-027-632-29552
; Sequence 29552, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29551
; LENGTH: 849
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-29552

Query Match
Best Local Similarity 30.3%; Score 91; DB 15; Length 849;
Matches 121; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
```



```
/ FILE REFERENCE: 108827.129
/ CURRENT APPLICATION NUMBER: US/10/027,632
/ CURRENT FILING DATE: 2002-04-30
/ PRIOR APPLICATION NUMBER: US 60/218,006
/ PRIOR FILING DATE: 2000-07-12
/ PRIOR APPLICATION NUMBER: US 60/198,676
/ PRIOR FILING DATE: 2000-04-20
/ PRIOR APPLICATION NUMBER: US 60/193,483
/ PRIOR FILING DATE: 2000-03-29
/ PRIOR APPLICATION NUMBER: US 60/185,218
/ PRIOR FILING DATE: 2000-02-24
/ PRIOR APPLICATION NUMBER: US 60/167,363
/ PRIOR FILING DATE: 1999-11-23
/ PRIOR APPLICATION NUMBER: US 60/156,358
/ PRIOR FILING DATE: 1999-09-28
/ PRIOR APPLICATION NUMBER: US 60/146,002
/ PRIOR FILING DATE: 1999-08-09
/ NUMBER OF SEQ ID NOS: 325720
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 29552
/ LENGTH: 849
/ TYPE: DNA
/ ORGANISM: Human
US-10-027-632-29552
```

Query Match 30.3%; Score 91; DB 15; Length 849;

Best Local Similarity 70.8%; Pred. No. 3.2e-13; Mismatches 50; Indels 0; Gaps 0;

```
QY 4 GGCCTGTATCCAGCTACTACAGAGGCTGAGCGAGAGGATCCGCGAGCTTGAGAGAT 63
DB 410 GGCTTGAAGTCCAGGTAAGTGAAGAGCTGAGGAGAGATCGCTTGAGCCAGAGAGA 469
QY 64 CTGCTTGAAGCTGAGAGGTTGAGGCTACAGTAAGCCAAATCATGCCATATCTTACG 123
DB 470 TCGCTTGAAGCCAGAGAGTTCACGCTTGACAGTGAAGCTATGATTTGGCCACTGATCCAGC 529
QY 124 CTGGGCGCAACAAGTGAAGCCCTAACAAAAAATTTAAAAAGAA 174
DB 530 CTGGGTGACAGAGCAAGCTTTGTCCCAAAAAAAGAAAAAGAAA 580
```

## RESULT 14

```
US-09-782-378A-17/c
/ Sequence 17, Application US/09782378A
/ Patent No. US20020102731A1
/ GENERAL INFORMATION:
/ APPLICANT: Hearing, Patrick
/ APPLICANT: Bahou, Wadie
/ APPLICANT: Sandalon, Ziv
/ APPLICANT: Gnatenko, Dmitri
/ TITLE OF INVENTION: Adenoviral Vectors
/ FILE REFERENCE: STONY-04970
/ CURRENT APPLICATION NUMBER: US/09/782,378A
/ CURRENT FILING DATE: 2001-02-12
/ PRIOR APPLICATION NUMBER: 60/237,747
/ PRIOR FILING DATE: 2000-10-02
/ NUMBER OF SEQ ID NOS: 27
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 17
/ LENGTH: 56737
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-09-782-378A-17
```

Query Match 30.3%; Score 91; DB 9; Length 56737;

Best Local Similarity 65.9%; Pred. No. 1.5e-12; Mismatches 75; Indels 1; Gaps 1;

```
QY 1 GGGGGCTGTATCCAGCTACTCAGAGGCTGAGAGGAGATCCGCGAGGCTGGC- 59
DB 3504 GCGCACTGTATGTCAGCTACTTGGAGGCTGAAGTGGAGAAATCCCTGAGGCTGGG 3445
```

```
QY 60 AGATCTGCTTACGCTTGGAGGCTTGAAGCTACAGTAAGCCAAATCATGCGAGTAATT 119
DB 3444 AGAATCAACCCAGGCGCGGAAGTGAAGCTGAGGAGTGAAGTGTGCGACAGCTC 3385
QY 120 CAGCTGGGCGCAAGTGAAGCCGTAACAAAAAAATTTAAAAAAGAAATTTA 179
DB 3384 CATCTAGGTACAGAGTGAAGCCCTGCTTCAAAAAAAGAAATTTGCGAGAAATTAAGTAA 3325
QY 180 GATCAAGATCCAACTGTAAAAAGTGCCCTAAACCAACATTTAA 222
DB 3324 GTTATGTTTGAAGATGAATAAATCAATTTTTCTCAGGAA 3282
```

## RESULT 15

```
US-10-027-632-223820/c
/ Sequence 223820, Application US/10027632
/ Publication No. US20030204075A9
/ GENERAL INFORMATION:
/ APPLICANT: Wang, David G.
/ TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
/ FILE REFERENCE: 108827.129
/ CURRENT APPLICATION NUMBER: US/10/027,632
/ CURRENT FILING DATE: 2002-04-30
/ PRIOR APPLICATION NUMBER: US 60/218,006
/ PRIOR FILING DATE: 2000-07-12
/ PRIOR APPLICATION NUMBER: US 60/198,676
/ PRIOR FILING DATE: 2000-04-20
/ PRIOR APPLICATION NUMBER: US 60/193,483
/ PRIOR FILING DATE: 2000-03-29
/ PRIOR APPLICATION NUMBER: US 60/185,218
/ PRIOR FILING DATE: 2000-02-24
/ PRIOR APPLICATION NUMBER: US 60/167,363
/ PRIOR FILING DATE: 1999-11-23
/ PRIOR APPLICATION NUMBER: US 60/156,358
/ PRIOR FILING DATE: 1999-09-28
/ PRIOR APPLICATION NUMBER: US 60/146,002
/ PRIOR FILING DATE: 1999-08-09
/ NUMBER OF SEQ ID NOS: 325720
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 223820
/ LENGTH: 595
/ TYPE: DNA
/ ORGANISM: Human
US-10-027-632-223820
```

Query Match 30.2%; Score 90.6; DB 15; Length 595;

Best Local Similarity 66.8%; Pred. No. 3.6e-13; Mismatches 64; Indels 0; Gaps 0;

```
QY 3 GGGCTGTATCCAGCTACTCAGAGGCTGAGCGAGAGGATCCGCGAGGCTGGGAGA 62
DB 288 GGCCTTAATGTTCCAGCTACTTGAAGGCTGAGGAGGATCATTTGAGCCTGGGAGG 229
QY 63 TCGCTGAGCTGAGAGGTTGAGGCTACAGTAAGCCAAATCATGCGAGTATCTTGA 122
DB 228 ATCATTTGAGCCAGAGAGTGAAGGCTCAAGTGAAGCTATGATGCGATCACTGCACTCAG 169
QY 123 CCTGGGCGCAAAAGTGAAGCCGTAACAAAAAAATTTAAAAAAGAAATTTAGAT 182
DB 168 CCTGGGTGACAGAGTGAAGCCCTGTCCAAAGCAACAAACAAAAAAGAAAGAA 109
QY 183 CAAGATCCAACTG 195
DB 108 ATCATTTGCAACAG 96
```

Search completed: March 13, 2004, 05:30:47  
Job time: 107.276 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd

OM nucleic - nucleic search, using sw model

```
Run on:      March 12, 2004, 13:19:34 ; Search time 955.276 Seconds
              (without alignments)
              9378.080 Million cell updates/sec
```

```

Title:      US-09-966-880A-10_COPY_1_300
Perfect score: 300
Sequence:   1  gggggcgttaaccagct.....aatgcatggtgcacctggt 300

```

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

```

1:  est:*
2:  em_estba:*
3:  em_esthun:*
4:  em_estin:*
5:  em_estmu:*
6:  em_estrov:*
7:  em_estrpl:*
8:  em_estro:*
9:  em_hnc:*
10:  gb_est1:*
11:  gb_est2:*
12:  gb_hnc:*
13:  gb_est3:*
14:  gb_est4:*
15:  gb_est5:*
16:  em_estfun:*
17:  em_ges_hum:*
18:  em_ges_inv:*
19:  em_ges_pln:*
20:  em_ges_vrt:*
21:  em_ges_fun:*
22:  em_ges_mam:*
23:  em_ges_mus:*
24:  em_ges_pro:*
25:  em_ges_rod:*
26:  em_ges_phg:*
27:  em_ges_vrl:*
28:  gb_gses1:*
29:  gb_gses2:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Length	DB	ID	Description
1	104.2	34.7	235	13	Bt665383
2	101.8	33.9	617	28	AQ383054
3	97.6	32.5	509	28	AQ315617
4	97	32.3	566	28	AQ31557
					PC1-11-3

C 5	95.2	31.7	526	12	BG221766	BG221766 RST41581
C 6	94.8	31.6	499	9	A1830009	A1830009 wj99604.x
C 7	94.8	31.6	495	28	AQ236867	AQ236867 RPTC11-71
C 8	94.8	31.6	666	28	B88182	B88182 RPTC11-2311
C 9	94.6	31.5	570	29	AG019309	AG019309 Homo sapti
C 10	94.6	31.5	679	29	AG019484	AG019484 Homo sapti
C 11	93.8	31.3	758	14	CA428264	CA428264 UI-H-DRD-
C 12	93	31.0	381	13	BQ311376	BQ311376 RCS-NM106
C 13	93	31.0	435	9	AA515045	AA515045 nrg67n05.s
C 14	92.6	30.9	748	10	BE742033	BE742033 601594512
C 15	92.4	30.8	276	10	BF916517	BF916517 IL3-UT011
C 16	92.4	30.8	660	28	AQ021972	AQ021972 CIT-HSP-2
C 17	92.2	30.7	314	9	AA828637	AA828637 od79d11.s
C 18	91.8	30.6	343	28	AQ053384	AQ053384 CIT-HSP-2
C 19	91.4	30.5	735	28	AQ0319539	AQ031953 RPTC11-92
C 20	88.8	29.6	367	28	AQ0319539	AQ031953 RPTC11-92
C 21	88.8	29.6	619	13	BX487715	BX487715 DXFZP686K
C 22	88.8	29.6	673	28	BZ608116	BZ608116 WHAC04TF
C 23	87.8	29.3	607	28	AO554450	AO554450 RPTC-11-4
C 24	87.4	29.1	319	9	AM088871	AM088871 xd05h02.x
C 25	87.4	29.1	745	13	BUE54170	BUE54170 AGENCOURT
C 26	87.4	29.1	936	28	AQ089972	AQ089972 HS-5147.B
C 27	86.4	28.8	566	28	AQ346062	AQ346062 RPTC11-12
C 28	86.4	28.8	679	13	BUE63021	BUE63021 UI-H-FBI-1
C 29	86.4	28.8	952	28	AO743668	AO743668 HS-5507.A
C 30	86.2	28.7	453	10	BE160516	BE160516 RCI-HT041
C 31	86.2	28.7	501	9	AA722505	AA722505 4h31g08.s
C 32	86.2	28.7	1051	10	BG026229	BG026229 602291607
C 33	86	28.7	452	9	AA668273	AA668273 ad78e07.s
C 34	86	28.7	845	14	CB993149	CB993149 AGENCOURT
C 35	85.2	28.4	371	9	A1148153	A1148153 qb56a08.x
C 36	85.2	28.4	787	28	BZ599553	BZ599553 WHADW27TF
C 37	84.6	28.2	768	28	BZ610312	BZ610312 WHAD102TF
C 38	84.4	28.1	314	9	A1251576	A1251576 qv45D02.x
C 39	84.4	28.1	872	10	BF680286	BF680286 602155052
C 40	84.4	28.1	896	13	BQ924330	BQ924330 AGENCOURT
C 41	84.2	28.1	177	10	AMW34029	AMW34029 QVQ-TT001
C 42	84.2	28.1	370	13	BX485214	BX485214 DXFZP686E
C 43	84.2	28.1	630	28	AG159351	AG159351 Pan tlogl
C 44	84.2	28.1	668	28	AG159359	AG159359 Pan tlogl
C 45	84	28.0	331	9	AL135357	AL135357 DXFZP762E

## ALIGNMENTS

RESULT 1	
B0665283	
LOCUS	235 bp
DEFINITION	B0665283 linear EST 30-SEP-2002
	c1129b12.z1 Hembase; Erythroid Precursor Cells (LB:c1 library)
	Homo sapiens cDNA clone c1129b12 5' , mRNA sequence.

JOURNAL COMMENT

Unpublished (2002)  
Contact: Jeffery L. Miller  
Laboratory of Chemical Biology  
National Institute of Diabetes and Digestive and Kidney Diseases  
Building 10, Room 9B17, National Institutes of Health, Bethesda, MD  
20892, USA  
Tel: 301 402 2373  
Fax: 301 435 5148  
Email: jlm7@nih.gov

The 'cl' library was constructed by Alexander Gubin, Ph.D. in the Laboratory of Chemical Biology, NIDDK, NIH. DNA Sequencing and/or analyses by National Institutes of Health Intramural Sequencing



Center (NISC). More information available at:

<http://hembase.niddk.nih.gov>

Plate: 129 row: b column: 12

Seq primer: 5' lambda-TripLex2 Sequencing Primer.

Location/Qualifiers

# FEATURES

Source

1. 235

/organism="Homo sapiens"

/mol\_type="RNA"

/db\_xref="taxon:9606"

/clone="cl129b12"

/sex="unknown"

/tissue\_type="blood"

/cell\_type="Erythroid Precursor Cells"

/cell\_line="Primary Culture of Peripheral Blood Mononuclear Cells"

/dev\_stage="Precursor erythroblasts; GFA++"

/lab\_host="DH5alpha"

/clone\_id="Hembase; Erythroid Precursor Cells (LCB:cl129b12)"

/note="Organ: blood; Vector: pTriplex2; Site 1: SfiI; Site 2: SfiI. A complementary DNA (cDNA) library from human erythroid precursor cells was constructed using SMART PCR (polymerase chain reaction) cDNA Library Construction Kit (Clontech, Palo Alto, CA) according to the manufacturer's directions, but with slight modifications. Briefly, reverse transcription was performed in the presence of 1 umol/l peptide nucleic acid (PNA) oligos

(N-terminal) -biotin-GTC-CAC-CCG-AGG-CTT-G-(C-terminal) and (N-terminal) -biotin-C(T/C)T-GAA-GTT-CTC-AGG-A-(C-terminal)

. Synthesized cDNA was digested with SfiI and site-selected on a 1% agarose gel (>800bp). Large-scale sequencing of the library was performed by the NIH Intramural Sequencing Center (NISC);

<http://www.nisc.nih.gov/>."

Query Match 34.7%; Score 104.2; DB 13; Length 235;

Best Local Similarity 75.1%; Pred. No. 1.5e-08;

Matches 130; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

# ORIGIN

Query Match 34.7%; Score 104.2; DB 13; Length 235;

Best Local Similarity 75.1%; Pred. No. 1.5e-08;

Matches 130; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

3 GGGCCTGATCCAGCTACTCAGAGAGCTGAGCAGAGATCCGGAGCCTGACAG 62

5 GGGCGGTATCCAGCTACTTGAAGAGCTGAGCAGAGATCTTGAACCCAGAGA 64

63 TCTGCTGAGCCTGGAGGTTGAGGCTPACGTAAGCCAGATCATGCAATATTCTT 122

65 ATTGCTTGAACCTGGAGGCGGAGGTTGCAATGAGCCAAAGATCTTGCAATTCAG 124

123 CCTGGGCGCAAAAGTGAAGCCGTAAACAAAAAATTTAAAGAAA 175

125 CCTGGGTGACAGATCGAGATCCGCTTAAAAAATTTAAAGAAA 177

RESULT 2  
LOCUS AQ383054/c 617 bp DNA linear GSS 21-MAY-1999

DEFINITION RPI11-139K16.TV RPI1-11 Homo sapiens genomic clone RPI1-11-139K16, genomic survey sequence.

ACCESSION AQ383054

VERSION AQ383054

KEYWORDS GSS.

ORGANISM Homo sapiens (human)

SOURCE Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 617)

Zhao, S., Adams, W.D., Nierman, W., Malek, J., de Jong, P. and Venter, J.C.

Use of BAC End Sequences from Library RPI1-11 for Sequence-Ready Map Building (1997)

Other\_GSSs: RPI11-139K16.TV

JOURNAL

COMMENT

Other\_GSSs: RPI11-139K16.TV

Use of BAC End Sequences from Library RPI1-11 for Sequence-Ready Map Building (1997)

Other\_GSSs: RPI11-139K16.TV

JOURNAL

COMMENT

Other\_GSSs: RPI11-139K16.TV

Contact: Shaying Zhao, William Nierman, Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850

Tel: 301 838 0200

Fax: 301 838 0208

Email: hbe@tigr.org

Clones are derived from the human BAC library RPI1-11. For BAC library availability, please contact Pieter de Jong

(pieterdejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from

Research Genetics ([info@resgen.com](http://info@resgen.com)). BAC end search page:

[http://www.tigr.org/cdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/cdb/humgen/bac_end_search/bac_end_search.html)

Seq primer: SP6

Class: BAC ends.

# FEATURES

Source

1. 617

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="GDB:7553247"

/clone="RPI1-11-139K16"

/sex="Male"

/cell\_type="Lymphocytes"

/clone\_id="RPI1-11"

/note="Vector: pBAC3.6; Site 1: EcoRI; Site 2: EcoRI; RPI11 Human Male BAC Library"

# ORIGIN

Query Match 33.9%; Score 101.8; DB 28; Length 617;

Best Local Similarity 69.2%; Pred. No. 2.7e-08;

Matches 139; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

1 GGGGCTGATCCAGCTACTCAGAGAGCTGAGGAGAGATCCGGAGCCTGGCA 60

490 GGGCAGCTGATCCAGCTACTCAGAGAGCTGAGGAGAGATCCCTGAAGCGAGA 431

61 GATGCTGAGCCTGGAGGTTGAGGCTACGTAAGCCAAAGATCATGATATTCTT 120

430 GATGCTTGAACCCAGAGAGCGAGGTGAGTAAGTGTAGATCAACCACTGACTCC 371

121 AGCCTGGCGCAAAAGTGAAGCCGTAAACAAAAAATTTAAAGAAATTTAG 180

370 AGCCTGGGTGAAGGTTGAGATCCATCTCAAAAAAATTTAAAGAAATTTAG 311

181 ATCAAGATCCAACTGTAAAA 201

310 AAGGAAAAAGAAAGTGAAGA 290

RESULT 3

LOCUS AQ35617/c 509 bp DNA linear GSS 11-JUL-1998

DEFINITION CIT-HSP-2320N18.TF CIT-HSP Homo sapiens genomic clone 2320N18, genomic survey sequence.

ACCESSION AQ35617

VERSION AQ35617.1 GI:3301714

KEYWORDS GSS.

ORGANISM Homo sapiens (human)

SOURCE Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 509)

Adams, W.D., Rounley, S.D., Zhao, S., Field, C.E., Bass, S., Linher, K., Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H., Simon, M. and Venter, J.C.

Use of a random BAC End Sequence Database for Sequence-Ready Map Building (1998)

Unpublished (1998)

Other\_GSSs: CIT-HSP-2320N18.TF

Contact: Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

Use of a random BAC End Sequence Database for Sequence-Ready Map Building (1998)

Unpublished (1998)

Other\_GSSs: CIT-HSP-2320N18.TF

Contact: Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

Use of a random BAC End Sequence Database for Sequence-Ready Map Building (1998)

Unpublished (1998)

Other\_GSSs: CIT-HSP-2320N18.TF

Contact: Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research



9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208

Email: mdamas@tigr.org  
Clones are available from Research Genetics (info@resgen.com). BAC  
end search page:  
http://www.tigr.org/cdb/hungen/bac\_end\_search/bac\_end\_search.html.  
Seq primer: M13 Reverse  
Class: BAC ends.

## FEATURES

source

Location/Qualifiers  
1..509  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/clone="2320N18"  
/sex="Male"  
/cell\_type="Sperm"  
/clone\_lib="CIT-HSP"  
/note="Vector: pBelobAC11; Site\_1: HindIII; Site\_2:  
HindIII"

## ORIGIN

Query Match 32.5%; Score 97.6; DB 28; Length 509;  
Best Local Similarity 70.7%; Pred. No. 1.5e-07;  
Matches 130; Conservative 0; Mismatches 54; Indels 0; Gaps 0;  
  
QY 5 GCCTGAATCCAGCTACTCAGAGGCTGAGGACGAGATCCGCGAGCTGGCAGATC 64  
|||||  
Db 343 GCTTGTGTCAGCTAGTCAAGAGGCTGAGGAGGAGATCATCTGAGCTGGAGGAT 284  
|||||  
QY 65 TGCCCTGAGCTGGAGGTTGAGGCTACAGTAAGCCAGATCATGCGACTATATCTCAGCC 124  
|||||  
Db 283 CATTTGAACCCGAGGAGTTGAGGCTGAGTGAAGTATACCTACATCCGCGAGCC 224  
|||||  
QY 125 TGGCGGACAAAGTGAAGCCGTACAAAATAAATTTAAAAAGAAATTTAGATCA 184  
|||||  
Db 223 TGGATGACAGAGAGAGACCTGTCTTAAAAAAGAAAAAATTTAGATGAGAT 164  
|||||  
QY 185 AGAT 188  
|||  
Db 163 AGTT 160

RESULT 4  
AC531557/c 526 bp DNA linear GSS 18-MAY-1999  
LOCUS  
DEFINITION  
RPCI-11-349018.TV RPCI-11 Homo sapiens genomic clone  
AC531557  
AC531557  
AC531557.1 GI:4839339  
KEYWORDS  
SOURCE  
ORGANISM  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 526)  
Zhao,S., Adams,M.D., Nieman,W., Malek,J., de Jong,P. and  
Venier,J.C.  
Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready  
Map Building  
Unpublished (1997)  
Other\_GSSes: RPCI-11-349018.TV  
Contact: Shaying Zhao, William Nieman, Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: hbe@tigr.org

Clones are derived from the human BAC library RPCI-11. For BAC  
library availability, please contact Pieter de Jong  
(pieter@dong.med.buffalo.edu). Clones may be purchased from  
BACpac Resources (http://bacpac.med.buffalo.edu/ordering) or from

Research Genet cs (info@resgen.com). BAC end search page:  
http://www.tigr.org/cdb/hungen/bac\_end\_search/bac\_end\_search.html.  
Seq primer: T7  
Class: BAC ends.

## FEATURES

source

Location/Qualifiers  
1..526  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="GDB:7633985"  
/db\_xref="taxon:9606"  
/clone="RPCI-11-349018"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/clone\_lib="RPCI-11"  
/note="Vector: pBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
RPCI11 Human Male BAC Library"

## ORIGIN

Query Match 32.3%; Score 97; DB 28; Length 526;  
Best Local Similarity 67.7%; Pred. No. 1.9e-07;  
Matches 136; Conservative 0; Mismatches 65; Indels 0; Gaps 0;  
  
QY 1 GGGGGCTGTAATCCAGCTACTCAGAGGCTGAGGACGAGATCCGCGAGCTGGCA 60  
|||||  
Db 488 GGGCACCCTGTATCCAGCTACTGAGGAGGCTTGAAGCAGATCGTTGAACGAGTA 429  
|||||  
QY 61 GATCTGCTGAGGCTGGAGGTTGAGGCTACAGTAAGCCAGATCATGCGACTATCTTC 120  
|||||  
Db 428 GAATCGCTGAACCCAGAGGCGAGAGGTTGCGATAGCTGAGATCAACCGCTGCACCTC 369  
|||||  
QY 121 AGCCTGGGCGACAAAGTGAAGCCGTACAAAATAAATTTAAAAAGAAATTTAG 180  
|||||  
Db 368 AGCCTGGGCGACAGGTTGAGGCTGAGTGCATCTCAAAAAAATAAAGAAATTAAGA 309  
|||||  
QY 181 ATCAAGATCCACTTAAAA 201  
|||  
Db 308 AAGAAAAAGAAAGTGAAGAGA 288

RESULT 5  
BG221766/c 596 bp mRNA linear EST 21-APR-2001  
LOCUS  
DEFINITION  
RST41581 Athereys RAGE Library Homo sapiens cDNA, mRNA sequence.  
BG221766  
BG221766.1 GI:13747787  
KEYWORDS  
SOURCE  
ORGANISM  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 596)  
Harrington,J.V., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,  
Cain,S., Leventhal,C., Thornton,M., Ramachandran,R.,  
Whittington,J., Lerner,R., Costanzo,D., McElligott,K., Bozzer,S.,  
Mays,R., Smith,E., Veloso,N., Klika,A., Hess,J., Colhoun,K., Lo,K.,  
Offenbacher,J., Danzig,J. and Ducar,M.  
Creation of genome-wide protein expression libraries using random  
activation of gene expression  
Nat. Biotechnol. 19 (5), 440-445 (2001)  
21227151  
11329013

JOURNAL  
MEDLINE  
PUBMED

Contact: Scott J. Cain  
Athersys, Inc.  
3201 Carnegie Ave, Cleveland, OH 44115, USA  
Tel: 216 431 9900  
Fax: 216 361 9596  
Email: scain@athersys.com

High quality sequence stop: 550.  
Location/Qualifiers  
1..596  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"

## FEATURES

source



/cell\_line="HT1080"  
/clone\_lib="Athenrys PAGE Library"  
/note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."

## ORIGIN

Query Match 31.7%; Score 95.2; DB 12; Length 596;  
Best Local Similarity 72.1%; Pred. No. 3.7e-07;  
Matches 124; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 1 GGGGCGCTGTAATCCAGTACTCTCAGAGGCTGAGGCGAGAGATCCGCGAGGCTGCA 60  
DB 286 GGGGCGCTGTAATCCAGTACTCTCAGAGGCTGAGGCGAGAGATCCGCGAGGCGAGG 227  
QY 61 GATCTGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTG 120  
DB 226 AATAGGCGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 167  
QY 121 AGGCTGCGGCGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 172  
DB 166 AGGCTGCGGCGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 115

RESULT 6 A1832009 422 bp mRNA linear EST 21-DEC-1999  
A1832009 wj99a04.x1 NCI CGAP Lym12 Homo sapiens CDNA clone IMAGE:2410926 3'  
DEFINITION similar to contains Alu repetitive element; contains element MER22  
repetitive element; mRNA sequence.  
A1832009 A1832009.1 GI:5452680  
EST.  
VERSION A1832009.1 GI:5452680  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE Mammalia; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 422)  
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapsb-rc@mail.nih.gov  
Life Technologies catalog #: 11547-015  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CCAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/dbp/image/image.html  
Insert Length: 538 Std Error: 0.00  
Seq primer: -40UP from Gibco  
High quality sequence stop: 404.  
Location/Qualifiers

FEATURES  
source  
1..422  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2410926"  
/tissue\_type="Lymphoma, follicular mixed small and large  
cell"  
/lab\_host="DH10B"  
/clone\_lib="NCI CGAP Lym12"  
/note="Organ: Lymph node; Vector: PCMV-SpOrE; Site 1:  
SalI; Site 2: NotI; Cloned unidirectionally. Primer:  
Oligo dt. Average insert size 1.25 kb. Life Technologies  
catalog #: 11547-015"

## ORIGIN

Query Match 31.6%; Score 94.8; DB 9; Length 422;  
Best Local Similarity 72.4%; Pred. No. 5e-07;

Matches 123; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 5 GCCTGTAATCCAGTACTCTCAGAGGCTGAGGCGAGAGATCCGCGAGGCTGAGAGATC 64  
DB 152 GCTTGAGTCCAGTACTCTCAGAGGCTGAGGCGAGAGATCCGCGAGGCTGAGAGAT 211  
QY 65 TGCTTGAAGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 124  
DB 212 TGCTTGAAGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 271  
QY 125 TGGGCGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 174  
DB 272 TGGGCGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 321

RESULT 7 A0236857/c 495 bp DNA linear GSS 21-APR-1999  
A0236857/c RPI11-71K16.TU RPI1-11 Homo sapiens genomic clone RPI1-11-71K16,  
genomic survey sequence.  
DEFINITION RPI11-71K16.TU RPI1-11 Homo sapiens genomic clone RPI1-11-71K16,  
genomic survey sequence.  
A0236857 A0236857.1 GI:3669148  
GSS.  
VERSION A0236857.1 GI:3669148  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE Mammalia; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 495)  
Adams, M.D., Rounsley, S.D., Zhao, S., Bass, S., Linher, K., Golden, K.,  
Berry, K., Granger, D., Suh, E., Wible, C., de Jong, P., and Venter, J.C.  
Use of human BAC End Sequences for Sequence-Ready Map Building  
Unpublished (1998)  
JOURNAL Contact: Mark Adams  
COMMENT Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdadams@tigr.org  
Clones are derived from the human BAC library RPI1-11. For BAC  
library availability, please contact Pieter de Jong  
(pieder@jeong.med.bufileo.edu). Clones may be purchased from  
BACPAC Resources (http://bacpac.med.bufileo.edu/ordering) or from  
Research Genetics (info@resgen.com). BAC end search page:  
http://www.tigr.org/cdb/humgen/bac\_end\_search/bac\_end\_search.html  
Seq primer: Sp6  
Class: BAC ends.  
Location/Qualifiers

## FEATURES

source  
1..495  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="GDB:7527135"  
/db\_xref="taxon:9606"  
/clone="RPI1-11-71K16"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/clone\_lib="RPI1-11"  
/note="Vector: pBAC3.6; Site 1: EcoRI; Site 2: EcoRI;  
RPI11 Human Male BAC Library"

## ORIGIN

Query Match 31.6%; Score 94.8; DB 28; Length 495;  
Best Local Similarity 65.7%; Pred. No. 4.7e-07;  
Matches 138; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 1 GGGGCGCTGTAATCCAGTACTCTCAGAGGCTGAGGCGAGAGATCCGCGAGGCTGCA 60  
DB 251 GGGGCGCTGTAATCCAGTACTCTCAGAGGCTGAGGCGAGAGATCCGCGAGGCTGCA 192  
QY 61 GATCTGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 120  
DB 191 GAATCATTTGAACCCAGAGAGAGGTTGCAATGAGCCAAAGTATGCCCATGTAATTC 132



QY	121	AGCGTGGCGCAAAAGTGAAGACCGTAACAAAATAATTTAAATAAATAATTAG	180
Db	131	AGCGTGGCGCAACAGTGAAGTCTGTCTCAAGAAAAATAAATAAACCACAAACATTAT	72
QY	181	ATCAAGATCCACTGTATAAAAGTGCCCTTA	210
Db	71	TCTGCAATATGAAGCAAGTGACTCA	42
RESULT 8			
B88182/c			
LOCUS			
DEFINITION	B88182	666 bp	DNA
ACCESSION	RP011-23114.7PB	RP01-11	Homo sapiens genomic clone
VERSION	B88182		
KEYWORDS	B88182.1	GI:2929314	
ORGANISM	GSS.		
SOURCE	Homo sapiens (human)		
REFERENCE	Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Adams, M.D., Rounsailey, S.D., Zhao, S., Field, C.E., Bass, S., Linher, K., and Golden, K., Bourne, J., Granger, D., Suh, E., Wible, C., de Jong, P., and Venter, J.C.		
TITLE	Use of BAC End Sequences for Sequence-Ready Map Building (1998)		
JOURNAL	Unpublished (1998)		
COMMENT	Contact: Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: mdadams@tigr.org Clones are derived from the human BAC library RP01-11. For BAC library availability, please contact Pieter de Jong (pieter@tigr.org, med.bufileo.edu). Clones may be purchased from BACDAC Resources ( <a href="http://bacpac.med.bufileo.edu/ordering">http://bacpac.med.bufileo.edu/ordering</a> ) or from Research Genetics ( <a href="http://www.rg.org/cdb/hungem/bac_end_search/bac_end_search.html">http://www.rg.org/cdb/hungem/bac_end_search/bac_end_search.html</a> ) Seq primer: SP6 Class: BAC ends.		
FEATURES			
source	location/Qualifiers		
	1..666		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="GDB:7508725"		
	/db_xref="taxon:9606"		
	/clone="RP01-11-23114"		
	/sex="Male"		
	/cell_type="lymphocytes"		
	/clone_1ib="RP01-11"		
	/note="Vector: pBACes3.6; Site 1: EcoRI; Site 2: EcoRI; RP011 Human Male BAC Library"		
ORIGIN			
Query Match	31.6%;	Score 94.8;	DB 28;
Best Local Similarity	65.7%;	Pred. No. 4.2e-07;	
Matches 138;	Conservative 0;	Mismatches 72;	Indels 0;
	Gaps 0		
QY	1	GGGGGCGCTGTATATCCAGTACTCTCAGAGGCGTGAAGCAGAGAGATTCGGGAGCGCTGCGA	60
Db	210	GGGTGCTGTATATCTCAGTACTCTCAGAGGCGTGAAGCAGAGAGATTCATTGGACGACGA	151
QY	61	GATCGCCCGTGAAGCGGAGGAGTGAAGGCGTGAAGCAGAGATTCAGTATCTTC	120
Db	150	GATATATTGAACCCAGAGGCGAGAGGTTGCGATGAGCCAAAGTTGCCCACTGATTTTC	91
QY	121	AGCTTGGCGCAAAAGTGAAGACCGTAACAAAATAATTTAAATAAATAATTAG	180
Db	90	AGCTTGGCGCAACAGTGAAGTCTGTCTCAAGAAAAATAAATAAACCACAAACATTAT	31
QY	181	ATCAAGATCCACTGTATAAAAGTGCCCTTA	210

	DB	30	TGCTGCATATGAAGCAGAAGTGTCATCA	1
	RESULT 9			
	LOCUS	AG019309/c	570 bp	DNA linear GSS 09-JAN-2003
	DEFINITION	Homo sapiens genomic DNA, 21q region, clone: B102017 A044(-21),		
	ACCESSION	AG019309		
	VERSION	AG019309.1 GI:6045253		
	KEYWORDS	GSS.		
	SOURCE	Homo sapiens (human)		
	ORGANISM	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
	REFERENCE			
	AUTHORS	Hattori,M., Fujiyama,A., Ishii,K., Toyoda,A., Taylor,T., Park,H.-S., Yada,T., Watanabe,H. and Sakaki,Y. Published Only in Database (1999) 2 (bases 1 to 570) Homo sapiens genomic DNA, chromosome 21q Direct Submission Submitted (16-OCT-1999) Masahira Hattori, RIKEN Genomic Sciences Center(GSC) c/o Kitasato University, 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan (E-mail:hattori@ngc.ims.u-tokyo.ac.jp, Tel:+81-42-778-9923, Fax:-42-778-9924) Location/Qualifiers 1..570 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606" /chromosome="21" /map="21q" /clone="B102017 A044(-21)"		
	FEATURES			
	SOURCE			
	ORIGIN			
	Query Match	31.5%	Score 94.6;	DB 29; Length 570;
	Best Local Similarity	75.2%;	Pred.No. 4.8e-07;	
	Matches 118;	Conservative 0;	Mismatches 39;	Indels 0; Gaps 0;
CY	5	GCCGTGTAATCCCAGTACTACTACAGAGGGCGTAGAGGCCAGAGCATCCCGGAGCCTGGACAATC	64	
Db	266	GCCGTGTAATCCCAGTACTACTACTAGAGGGCGTAGAGTGAAGGAAATCCCTTGAAGCCCCGGAATGTT	207	
CY	65	TGCCGTGAGCCCTGGAGAGGTGAGGCTACAGTAAAGCAGCAAGATCATGCAGTATTACTTCAGACC	124	
Db	206	CGCTTCAGCCCTGGAGTGTGAGGCTGAGGCGTCAATTTGTGTCACTGCATGCACTCCAGACC	147	
CY	125	TGGGCGAACAAAGTAGAAGCCGTAAACAAAAAAAATAA	161	
Db	146	TGGGTGCCAGAGTAGAGACCCGTCTCAAAAAAAAAAA	110	
RESULT 10				
LOCUS	AG019484/c	679 bp	DNA linear GSS 09-JAN-2003	
DEFINITION	Homo sapiens genomic DNA, 21q region, clone: B102017_N009_Fw,			
ACCESSION	AG019484			
VERSION	AG019484.1 GI:6045428			
KEYWORDS	GSS.			
SOURCE	Homo sapiens (human)			
ORGANISM	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE				
AUTHORS	Hattori,M., Fujiyama,A., Ishii,K., Toyoda,A., Taylor,T., Park,H.-S., Yada,T., Watanabe,H. and Sakaki,Y. Published Only in Database (1999)			
TITLE	JOURNAL			



REFERENCE 2 (bases 1 to 679)  
AUTHORS Hattori, M., Fujiyama, A., Ishii, K., Toyoda, A., Taylor, T.,  
Park, H.-S., Yada, T., Matanabe, H. and Sakaki, Y.  
TITLE Direct Submission  
JOURNAL Submitted (16-OCT-1999) Masahira, Hattori, RIKEN Genomic Sciences  
Center (GSC) c/o Kitasato University, 1-15-1 Kitasato, Sagamihara,  
Kanagawa 228-8555, Japan (E-mail: hattori@gsc.ims.u-tokyo.ac.jp,  
Tel: 81-42-778-9923, Fax: 81-42-778-9924)

FEATURES  
source Location/Qualifiers  
1..679  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="21"  
/map="21q"  
/clone="B102017\_N009\_Fw"

## ORIGIN

Query Match 31.5%; Score 94.6; DB 29; Length 679;  
Best Local Similarity 75.2%; Pred. No. 4.5e-07;  
Matches 118; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 5 GCCTGTAATCCAGCTACTCTACAGAGGCTGAGGCGAGAGATCCGGAGCCTGGCAATC 64  
DB 421 GCCTGTAATCCAGCTACTCTACAGAGGCTGAGGCGAGAGATCCGGAGCCTGGCAATC 362  
QY 65 TGGCTGAGCTGAGCTGAGGCTGAGGCTGAGGCGAGAGATCCGGAGCCTGGCAATC 124  
DB 361 GCCTGAGCTGAGCTGAGGCTGAGGCTGAGGCGAGAGATCCGGAGCCTGGCAATC 302  
QY 125 TGGGCGACAAAGTGAACCGTACCAAAAAAAAAAAAAA 161  
DB 301 TGGGCGACAAAGTGAACCGTACCAAAAAAAAAAAAAA 265

## RESULT 11

CA428264 758 bp mRNA linear EST 07-NOV-2002

LOCUS UI-H-DFO-ber-o-01-0-UI.s1 NCI CGAP\_DFO Homo sapiens cDNA clone  
DEFINITION UI-H-DFO-ber-o-01-0-UI 3', mRNA sequence.  
ACCESSION CA428264  
VERSION CA428264.1 GI:24790990  
KEYWORDS EST.

SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
Tumor Gene Index  
Unpublished (1997)  
NATIONAL CANCER INSTITUTE, Cancer Genome Anatomy Project (CGAP).

JOURNAL CONTACT: Robert Strausberg, Ph.D.  
COMMENT Email: cga@bcr-remail.nih.gov

Tissue Procurement: Dr. Jose Mercuende  
cDNA library preparation: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Clone distribution information can be obtained  
from Dr. M. Bento Soares, bento-soares@iowa.edu  
The following repetitive elements were found in this cDNA  
sequence: 11-240, >ALU (matched complement) 416-674, >ALU (matched  
complement)  
Seq primer: M13 FORWARD  
POLYA=Yes.

FEATURES  
source Location/Qualifiers  
1..758  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-H-DFO-ber-o-01-0-UI"  
/issue\_type="Subchondral Bone"  
/dev\_stage="Adult"

/lab\_host="DH10B (Life Technologies)"

/clone\_lib="NCI CGAP DFO"

/note="Organ: Bone; Vector: pRT73-Pac (Pharmacia) with a  
modified polylinker; Site 1: EcoR I; Site 2: Not I;  
NCI CGAP DFO is a cDNA library containing the following  
tissue(s): Subchondral Bone. The library was constructed  
according to Bonaldi, Lennon and Soares, Genome Research,  
6:791-806, 1996. First strand cDNA synthesis was primed  
with an oligo-dT primer containing a Not I site. Double  
stranded cDNA was ligated to an EcoR I adaptor, digested  
with Not I, and cloned directionally into pRT73-Pac  
vector. The oligonucleotide used to prime the synthesis of  
first-strand cDNA contains a library tag sequence that is  
located between the Not I site and the (dT)18 tail. The  
sequence tag for this library is GTTAGCGTC.

TAG TISSUE=Subchondral bone  
TAG LIB=UI-H-DFO  
TAG\_SEQ=GTTAGCGTC"

## ORIGIN

Query Match 31.3%; Score 93.8; DB 14; Length 758;  
Best Local Similarity 75.7%; Pred. No. 5.9e-07;  
Matches 131; Conservative 0; Mismatches 37; Indels 5; Gaps 1;

QY 1 GGGGGGCTGTAATCCAGCTACTCTACAGAGGCTGAGGCGAGAGATCCGGAGCCTG--- 57  
DB 173 GGGGCTGTAATCCAGCTACTCTACAGAGGCTGAGGCGAGAGATCCGGAGCCTGAGG 114  
QY 58 --GCAATGCTGCTGAGCTGAGAGGTTGAGGCTACAGTAAGCCCAAGTCAAGTAT 115  
DB 113 CAGGAGATCACTTGAACCTAGAGGCGAGAGGTTGAGGCGAGAGATCCGGAGCCTGTC 54  
QY 116 ACTTCAGCCTGGGCGACAAAGTGAACCGTACCAAAAAAAAAATTAA 168  
DB 53 ACTTCAGCCTGGGCGACAAAGTGAACCGTACCAAAAAAAAAATTAA 1

## RESULT 12

BQ341376 361 bp mRNA linear EST 20-MAY-2002

LOCUS BQ341376  
DEFINITION BQ341376  
ACCESSION BQ341376  
VERSION BQ341376.1 GI:21003349  
KEYWORDS EST.

SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE Dias Neto, E., Garcia Correia, R., Verjovski-Almeida, S., Brites, M.R.,  
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, P.F.,  
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,  
Brunstein, A., deOliveira, P.S., Bucher, P., Jungeneel, C.V.,  
O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
PUBMED 10737800

COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil

Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(<http://www.ludwig.org.br/scripts/gethtml2.pl?l=RCS&t2=RCS-NN1065-300600-022-C04&t3=2000-06-30&t4=1>)  
Seq primer: puc 16 forward







insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

## ORIGIN

Query Match 30.9%; Score 92.6; DB 10; Length 748;  
Best Local Similarity 67.4%; Pred. No. 9.6e-07;  
Matches 145; Conservative 0; Mismatches 69; Indels 1; Gaps 1;

QY 1 GGGGCGCTGTAATCCAGCTACTCAGAGAGCTGAGGAGATCCGCGAGCTGGCA 60  
DB 244 GGGTCCCTGTAGTCCAGCTACTTGGAGGCTGAGGAGATGCGATGAGGCAAG- 186  
QY 61 GATCTGCTGAGCTTGGAGGTTGAGGTACAGTAACCAAGATCAGCCAGTACTTC 120  
DB 185 GAATGGCATGAACCTGGAGGTTGAGGTGAGGAGGATGGCACTGCACTCC 126  
QY 121 AGCGTGGGCGACAAGTGAGACCGTAACAAAAAATTTAAAAAAAAGAAATTAG 180  
DB 125 AGCTGGGCGACAGAGTGAAGTCTCTTTCCAAAAAATTTAACTCGAGACTAG 66  
QY 181 ATCAAGATCCAACTGTAAAAAAGTGCGCTTAAACACC 215  
DB 65 TTCTCTCTCTCTCTCTCTGACAGCGAGGCTCGTACC 31

RESULT 15  
BF916517 276 bp mRNA linear EST 18-JAN-2001  
LOCUS IL3-UT0115-131200-386-A03 UT0115 Homo sapiens cDNA, mRNA sequence.  
DEFINITION  
BF916517  
ACCESSION  
BF916517.1 GI:12308079  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS  
1 (bases 1 to 276)  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
Nagai, M.A., da Silva, M. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
Goldani, G.H., Carvalho, A.F., Matukuma, A., Bala, G.S., Simpson, D.H.,  
Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V.,  
O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
Simpson, A.J.

TITLE  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL  
MEDLINE  
PUBMED  
COMMENT  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?fil=IL3&t2=IL3-UT0115-  
131200-386-A03&t3=2000-12-13&t4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 276.

FEATURES  
source  
1..276  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/dev\_stage="Adult"  
/clone\_lib="UT0115"  
/note="Organ: uterus; tumor; Vector: puc18; Site: 1: SmaI;  
Site\_2: SmaI; A mini-library was made by cloning products

derived from ORFESTS PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

## ORIGIN

Query Match 30.8%; Score 92.4; DB 10; Length 276;  
Best Local Similarity 75.3%; Pred. No. 1.6e-06;  
Matches 128; Conservative 0; Mismatches 41; Indels 1; Gaps 1;

QY 1 GGGGCGCTGTAATCCAGCTACTCAGAGGCTGAGGAGATCCGCGAGCTGGCA 60  
DB 87 GGGTCCCTGTAGTCCAGCTACTTGGAGGCTGAGGAGATGCGATGAGGCAAG- 145  
QY 61 GATCTGCTGAGCTTGGAGGTTGAGGTACAGTAACCAAGATCAGCCAGTACTTC 120  
DB 146 GAATGGCATGAACCTGGAGGTTGAGGTGAGGAGGATGGCACTGCACTCC 205  
QY 121 AGCGTGGGCGACAAGTGAGACCGTAACAAAAAATTTAAAAAAA 170  
DB 206 AGCTGGGCGACAGAGTGAAGTCTCTCTCAAAAAAATTTAAACAAA 255

Search completed: March 13, 2004, 00:38:52  
Job time : 959.401 secs







FEATURES  
source

Location/Qualifiers  
1. .693  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:436722"  
/tissue\_type="primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_1lb="NIH\_MGC\_48"  
/note="Organ: B-cells; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 39.3%; Score 118.2; DB 10; Length 693;  
Best Local Similarity 97.6%; Pred. No. 1.2e-11;  
Matches 120; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 18 TTGAAGATTATTTTACTGCTGGAATCTTTGTAGAAAACCAAGAACTTCAAA 77  
Db 197 TTCAAGATTATTTTACTGCTGGAATCTTTGTAGAAAACCAAGAACTTCAAA 256  
QY 78 GCCTGGAGAGGCTGCATGAAATTCAGTTCCTCCAGACAGCTTCGGCCATCCTT 137  
Db 257 GCCTGGAGAGGCTGCATGAAATTCAGTTCCTCCAGACAGCTTCGGCCATCCTT 316  
QY 138 TTG 140  
Db 317 TTG 319

## RESULT 2

Bg686133

LOCUS 743 bp mRNA linear EST 01-MAY-2001  
DEFINITION 602638412F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4766234 5',  
mRNA sequence.

ACCESSION

Bg686133

Bg686133.1 GI:13917530

EST.

Homo sapiens

Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 743)

NIH-MGC http://mgs.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgaps-remail.nih.gov

Tissue Procurement: Louis M. Staudt, M.D., Ph.D.

CDNA Library Preparation: Ling Hong/Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

http://image.lnl.gov

Plate: LNCM1626 row: 5 column: 03

High quality sequence stop: 740.

Location/Qualifiers

1. .743

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:4766234"

/tissue\_type="primary B-cells from tonsils (cell line)"

/lab\_host="DH10B (phage-resistant)"

/clone\_1lb="NIH\_MGC\_48"

## ORIGIN

Query Match 39.3%; Score 118.2; DB 12; Length 743;  
Best Local Similarity 97.6%; Pred. No. 1.2e-11;  
Matches 120; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 18 TTGAAGATTATTTTACTGCTGGAATCTTTGTAGAAAACCAAGAACTTCAAA 77  
Db 498 TTCAAGATTATTTTACTGCTGGAATCTTTGTAGAAAACCAAGAACTTCAAA 557  
QY 78 GCCTGGAGAGGCTGCATGAAATTCAGTTCCTCCAGACAGCTTCGGCCATCCTT 137  
Db 558 GCCTGGAGAGGCTGCATGAAATTCAGTTCCTCCAGACAGCTTCGGCCATCCTT 617  
QY 138 TTG 140  
Db 618 TTG 620

## RESULT 3

Bg757089

LOCUS 820 bp mRNA linear EST 15-MAY-2001  
DEFINITION 602715124F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:485517 5',  
mRNA sequence.

ACCESSION

Bg757089

Bg757089.1 GI:14067742

EST.

Homo sapiens

Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 820)

NIH-MGC http://mgs.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgaps-remail.nih.gov

Tissue Procurement: Louis M. Staudt, M.D., Ph.D.

CDNA Library Preparation: Ling Hong/Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

http://image.lnl.gov

Plate: LNCM1704 row: 0 column: 06

High quality sequence stop: 675.

Location/Qualifiers

1. .820

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:485517"

/tissue\_type="primary B-cells from tonsils (cell line)"

/lab\_host="DH10B (phage-resistant)"

/clone\_1lb="NIH\_MGC\_48"

/note="Organ: B-cells; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(G). Size-selected &gt;500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."



## ORIGIN

Query Match 39.3%; Score 118.2; DB 12; Length 820;  
 Best Local Similarity 97.6%; Pred. No. 1.1e-11;  
 Matches 120; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 18 TTGAAGATTATTTTACTGCTGGAATCTTTTGTAGAAAACGACGAAGACTTTCAAA 77  
 |||  
 DB 505 TTCAAAGATTATTTTACTGCTGGAATCTTTTGTAGAAAACGACGAAGACTTTCAAA 564  
 |||  
 QY 78 GCGTGGGAAGGGCTGCATGAAATTCAGTTCGCTCCAGACAGCTTCGGCGCATCCTT 137  
 |||  
 DB 565 GCGTGGGAAGGGCTGCATGAAATTCAGTTCGCTCCAGACAGCTTCGGCGCATCCTT 624  
 |||  
 QY 138 TTG 140  
 |||  
 DB 625 TTG 627

## RESULT 4

BG758510 872 bp mRNA linear EST 15-MAY-2001  
 LOCUS 602712721 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4853069 5',  
 DEFINITION mRNA sequence.

ACCESSION BG758510  
 VERSION BG758510.1 GI:14069163  
 KEYWORDS EST.

SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 872)  
 NIH-MGC http://mgi.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: LNCM1698 row: 1 column: 06  
 High quality sequence step: 836.

## FEATURES

source

1..872  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4853069"  
 /tissue\_type="Primary B-cells from tonsils (cell line)"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_11b="NIH\_MGC\_48"  
 /note="Organ: B-cells; Vector: pOTB7; Site\_1: XhoI;  
 Site\_2: EcoRI; cDNA made by oligo-dT priming.  
 Directionally cloned into EcoRI/XhoI sites using the  
 following 5' adaptor: GGCACGAG(G). Size-selected >500bp  
 for average insert size 1.8kb. Library constructed by Ling  
 Hong in the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies).  
 Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 39.3%; Score 118.2; DB 12; Length 872;  
 Best Local Similarity 97.6%; Pred. No. 1.1e-11;  
 Matches 120; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 18 TTGAAGATTATTTTACTGCTGGAATCTTTTGTAGAAAACGACGAAGACTTTCAAA 77  
 |||  
 DB 486 TTCAAAGATTATTTTACTGCTGGAATCTTTTGTAGAAAACGACGAAGACTTTCAAA 545  
 |||

QY 78 GCGTGGGAAGGGCTGCATGAAATTCAGTTCGCTCCAGACAGCTTCGGCGCATCCTT 137  
 |||  
 DB 546 GCGTGGGAAGGGCTGCATGAAATTCAGTTCGCTCCAGACAGCTTCGGCGCATCCTT 605  
 |||  
 QY 138 TTG 140  
 |||  
 DB 606 TTG 608

## RESULT 5

BQ065440 953 bp mRNA linear EST 02-APR-2002  
 LOCUS AGENCOURT 6855061 NIH\_MGC\_99 Homo sapiens cDNA clone IMAGE:5929977  
 DEFINITION 5', mRNA sequence.

ACCESSION BQ065440  
 VERSION BQ065440.1 GI:19894486  
 KEYWORDS EST.

SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 953)  
 NIH-MGC http://mgi.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: Lou Staudt  
 cDNA Library Preparation: Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: LNCM2108 row: P column: 10  
 High quality sequence step: 634.

## FEATURES

source

1..953  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5929977"  
 /tissue\_type="Lymphoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_11b="NIH\_MGC\_99"  
 /note="Organ: lymph; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
 into EcoRI/XhoI sites using the following 5' adaptor:  
 GGCACGAG(G). Size-selected >500bp for average insert size  
 1.8kb. Library constructed by Ling Hong in the laboratory  
 of Gerald M. Rubin (University of California, Berkeley)  
 using ZAP-cDNA synthesis kit (Stratagene) and Superscript  
 II RT (Life Technologies). Note: this is a NIH\_MGC  
 Library."

## ORIGIN

Query Match 39.3%; Score 118.2; DB 13; Length 953;  
 Best Local Similarity 97.6%; Pred. No. 1e-11;  
 Matches 120; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 18 TTGAAGATTATTTTACTGCTGGAATCTTTTGTAGAAAACGACGAAGACTTTCAAA 77  
 |||  
 DB 484 TTCAAAGATTATTTTACTGCTGGAATCTTTTGTAGAAAACGACGAAGACTTTCAAA 543  
 |||  
 QY 78 GCGTGGGAAGGGCTGCATGAAATTCAGTTCGCTCCAGACAGCTTCGGCGCATCCTT 137  
 |||  
 DB 544 GCGTGGGAAGGGCTGCATGAAATTCAGTTCGCTCCAGACAGCTTCGGCGCATCCTT 603  
 |||  
 QY 138 TTG 140  
 |||  
 DB 604 TTG 606



RESULT 6  
 LOCUS B0055935  
 DEFINITION B0055935 1052 bp mRNA linear EST 29-MAR-2002  
 AGENCOURT 6796291 NIH\_MGC\_99 Homo sapiens cDNA clone IMAGE:5808181  
 5', mRNA sequence.  
 ACCESSION B0055935  
 VERSION B0055935.1 GI:19815262  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 1052)  
 NIH-MGC <http://mgs.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
 Tissue Procurement: Lou Staudt  
 cDNA Library Preparation: Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
 Plate: LICM2051 Row: m Column: 14  
 High quality sequence stop: 665.  
 Location/Qualifiers  
 1..1052  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5808181"  
 /tissue\_type="lymphoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_11b="NIH\_MGC\_99"  
 /note="Organ: lymph; Vector: pOTB7; Site 1: XhoI; Site 2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
 into EcoRI/XhoI sites using the following 5' adaptor:  
 GGCACGAG(G). Size-selected >500bp for average insert size  
 1.8kb. Library constructed by Ling Hong in the laboratory  
 of Gerald M. Rubin (University of California, Berkeley)  
 using ZAP-cDNA synthesis kit (Stratagene) and Superscript  
 II RT (Life Technologies). Note: this is a NIH\_MGC  
 library."

ORIGIN  
 Query Match 39.3%; Score 118.2; DB 13; Length 1052;  
 Best Local Similarity 97.6%; Pred. No. 9.8e-12;  
 Matches 120; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 18 TTGAAGATTATTTTACTGCTGCAATCTTTGTAGAAACCAAGAAAGACTTTCAA 77  
 |||  
 DB 484 TTCAAGATTATTTTACTGCTGCAATCTTTGTAGAAACCAAGAAAGACTTTCAA 543  
 |||  
 QY 78 GCCTGGGAAGGGCTGCATGAAATTCAGTTCCTCTCCAGACAGCTTCGGGCACTCTT 137  
 |||  
 DB 544 GCCTGGGAAGGGCTGCATGAAATTCAGTTCCTCTCCAGACAGCTTCGGGCACTCTT 603  
 |||  
 QY 138 TTG 140  
 |||  
 DB 604 TTG 606

RESULT 7  
 LOCUS AL559877  
 DEFINITION AL559877 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA  
 clone CS0D003YB14 5-PRIME, mRNA sequence.  
 ACCESSION AL559877  
 VERSION AL559877.2 GI:31284008  
 KEYWORDS EST.

SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 1201)  
 NIH-MGC <http://mgs.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (2001)  
 COMMENT On Feb 15, 2001 this sequence version replaced gi:12905793.  
 Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: [sequef@genoscope.cns.fr](mailto:sequef@genoscope.cns.fr); Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 6672.r For  
 more information about this cluster, see  
<http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0D003DA070P1&cluster=6672.r>. Contact :  
 Feng Liang Email : [liang@life-tech.com](mailto:liang@life-tech.com) URL :  
<http://fulllength.invitrogen.com/> Invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID : CS0D003DA070P1.  
 Location/Qualifiers  
 1..1201  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0D003YB14"  
 /tissue\_type="B CELLS (RAMOS CELL LINE)"  
 /cell\_line="RAMOS CELL LINE"  
 /clone\_11b="Homo sapiens B CELLS (RAMOS CELL LINE)"  
 /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed  
 with a NotI-oligo(dT) primer. Five prime end enriched,  
 double-strand cDNA was digested with Not I and cloned into  
 the Not I and EcoRV sites of the pCMVSPORT 6 vector.  
 Library was not normalized."

ORIGIN  
 Query Match 39.3%; Score 118.2; DB 9; Length 1201;  
 Best Local Similarity 97.6%; Pred. No. 9.2e-12;  
 Matches 120; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 18 TTGAAGATTATTTTACTGCTGCAATCTTTGTAGAAACCAAGAAAGACTTTCAA 77  
 |||  
 DB 512 TTCAAGATTATTTTACTGCTGCAATCTTTGTAGAAACCAAGAAAGACTTTCAA 571  
 |||  
 QY 78 GCCTGGGAAGGGCTGCATGAAATTCAGTTCCTCTCCAGACAGCTTCGGGCACTCTT 137  
 |||  
 DB 572 GCCTGGGAAGGGCTGCATGAAATTCAGTTCCTCTCCAGACAGCTTCGGGCACTCTT 631  
 |||  
 QY 138 TTG 140  
 |||  
 DB 632 TTG 634

RESULT 8  
 LOCUS BF975166  
 DEFINITION BF975166 942 bp mRNA linear EST 22-JUN-2001  
 602244657F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:433639 5',  
 mRNA sequence.  
 ACCESSION BF975166  
 VERSION BF975166.1 GI:12342381  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 942)  
 NIH-MGC <http://mgs.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D.



CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.lnl.gov>  
Plate: LLCM1207 row: a column: 16  
High quality sequence stop: 707.  
Location/Qualifiers

## FEATURES

1..942  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:435639"  
/issue\_type="primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_48"  
/note="Organ: B-cells; Vector: pOTB7; Site\_1: XhoI;  
Site\_2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCAAGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by Ling  
Hong in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 39.1%; Score 117.6; DB 10; Length 942;  
Best Local Similarity 96.8%; Pred. No. 1.3e-11;  
Matches 120; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 18 TTGAAGATTATTTTACTGCTGGATCTTTGTAGAAACACGAAAGACTTCAAA 77  
DB 499 TTCAAAGATTATTTTACTGCTGGATCTTTGTAGAAACACGAAAGACTTCAAA 558

QY 78 GCCTGGAGAGGGCTGCATGAAATTCAGTTCCTCCAGACAGCTTGGCGCATCCTT 137  
DB 559 GCCTGGAGAGGGCTGCATGAAATTCAGTTCCTCCAGACAGCTTGGCGCATCCTT 618

QY 138 TTGG 141  
DB 619 TGGG 622

RESULT 9  
BG757392 693 bp mRNA linear EST 15-MAY-2001  
LOCUS 602711022F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4851580 5',  
DEFINITION mRNA sequence.  
ACCESSION BG757392  
VERSION BG757392.1 GI:14068045  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 693)  
NIH-MGC <http://imgc.ncl.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-r@mail.nih.gov](mailto:cgabbs-r@mail.nih.gov)  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.lnl.gov>  
Plate: LLCM1694 row: K column: 05  
High quality sequence stop: 693.  
Location/Qualifiers

## FEATURES

## source

1..693  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4851580"  
/issue\_type="primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_48"  
/note="Organ: B-cells; Vector: pOTB7; Site\_1: XhoI;  
Site\_2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCAAGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by Ling  
Hong in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 38.7%; Score 116.6; DB 12; Length 693;  
Best Local Similarity 96.7%; Pred. No. 2.3e-11;  
Matches 119; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 18 TTGAAGATTATTTTACTGCTGGATCTTTGTAGAAACACGAAAGACTTCAAA 77  
DB 502 TTCAAAGATTATTTTACTGCTGGATCTTTGTAGAAACACGAAAGACTTCAAA 561

QY 78 GCCTGGAGAGGGCTGCATGAAATTCAGTTCCTCCAGACAGCTTGGCGCATCCTT 137  
DB 562 GCCTGGAGAGGGCTGCATGAAATTCAGTTCCTCCAGACAGCTTGGCGCATCCTT 621

QY 138 TTG 140  
DB 622 TTG 624

RESULT 10  
BX464579 853 bp mRNA linear EST 22-MAY-2003  
LOCUS BX464579 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA  
DEFINITION clone CS0DG003YB14 5-PRIME, mRNA sequence.  
ACCESSION BX464579  
VERSION BX464579.1 GI:31031641  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 853)  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: [seqref@genoscope.cns.fr](mailto:seqref@genoscope.cns.fr), Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 6672.r for  
more information about this cluster, see  
<http://www.genoscope.cns.fr/>  
cgi-bin/cluster.cgi?seq=CS1D5001ZF10QPI&cluster=6672.r. Contact :  
Feng Liang Email : [filang@lifetech.com](mailto:filang@lifetech.com) URL :  
<http://fulllength.invitrogen.com/> Invitrogen Corporation 1600  
Paradise Avenue Genoscope sequence ID : CS1D5001ZF10QPI.  
Location/Qualifiers

## FEATURES

1..853  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DG003YB14"  
/issue\_type="B CELLS (RAMOS CELL LINE)"  
/cell\_line="RAMOS CELL LINE"  
/clone\_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"

## FEATURES



/note="Vector: PCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the PCMVSPORT 6 vector. Library was not normalized."

## ORIGIN

Query Match 38.7%; Score 116.6; DB 13; Length 853;  
Best Local Similarity 96.7%; Pred. No. 2.1e-11;  
Matches 119; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 18 TTGAAGATTATTTTACTGCTGGAATCTTTTGTAGAAACCAAGAAAGACTTTCAA 77  
DB 506 TTCAAGATTATTTTCTGCTGGAATCTTTTGTAGAAACCAAGAAAGACTTTCAA 565  
QY 78 GCGTGGGAGGCGTCATGAAATTCAGTTCCTCTCCAGACAGCTTCGCGCATCTT 137  
DB 566 GCGTGGGAGGCGTCATGAAATTCAGTTCCTCTCTCCAGACAGCTTCGCGCATCTT 625  
QY 138 TTG 140  
DB 626 TTG 628

RESULT 11 1034 bp mRNA linear EST 15-MAY-2001  
BG755005 6027115.11F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4851819 5',  
LOCUS mRNA sequence.  
ACCESSION BG755005  
VERSION BG755005.1 GI:14065658  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1034)  
AUTHORS NIH-MGC <http://mhc.nhl.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
Plate: LNCM1695 row: e column: 04  
High quality sequence stop: 622.

## FEATURES

source

Location/Qualifiers  
1..1034  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4851819"  
/tissue\_type="primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_id="NIH\_MGC\_48"  
/note="Organ: B-cells; Vector: pOTB7; Site\_1: XhoI;  
Site\_2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCAAGAG(G). Size-selected  
for average insert size 1.8kb. Library constructed by Ling  
Hong in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 38.1%; Score 114.8; DB 12; Length 1034;  
Best Local Similarity 91.0%; Pred. No. 3.9e-11;

Matches 122; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 7 TTTTCTTTTGTGAAGATTATTTTACTGCTGGAATCTTTGTAGAAACCAAGAA 66  
DB 27 TTTTCTTTTGTGAAGATTATTTTACTGCGGGAATCTTTGTAGAAACCAAGAA 86  
QY 67 GAACCTTCAAGCGCTGGAGGCGTCGATGAAATTCAGTTCCTCTCCAGACAGCTTC 126  
DB 87 GAACCTTCAAGCGGAGGCGTCGATGAAATTCAGTTCCTCTCCAGACAGCTTC 146  
QY 127 GCGCATCTCTTTG 140  
DB 147 GCGCATCTCTTTG 160

RESULT 12 442 bp DNA linear GSS 30-NOV-2001  
BH302559/c  
LOCUS CH230-100C8.TV CHORI-230 Segment 1 Rattus norvegicus genomic clone  
DEFINITION CH230-100C8, genomic survey sequence.  
ACCESSION BH302559  
VERSION BH302559.1 GI:17214967  
KEYWORDS GSS.  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM Rattus norvegicus  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE 1 (bases 1 to 442)  
AUTHORS Zhao, S., Shetty, J., Shatsman, S., Tsagaye, G., Geer, K.,  
Shvartsbeyn, A., Gebregiorgis, E., Overton, L., Russell, D., Chen, D.,  
Riggs, F., de Jong, P. and Fraser, C.M.  
TITLE Rat BAC End Sequences from Library CHORI-230 EcoRI segment  
JOURNAL Unpublished (1999)  
COMMENT Other GSSs: CH230-100C8.TV  
Contact: Shanying Zhao  
Department of Bukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: [szhao@tigr.org](mailto:szhao@tigr.org)  
Clones are derived from the rat BAC library CHORI-230  
(<http://www.chori.org/bacpac/rat230.htm>). For BAC library  
availability, please contact Pieter de Jong ([pdj@omni.net](mailto:pdj@omni.net)).  
Clones may be purchased from BACPAC Resources  
([http://www.chori.org/bacpac/or\\_exting\\_information.htm](http://www.chori.org/bacpac/or_exting_information.htm)). BAC end  
pages: [http://www.tigr.org/tdb/bac\\_end/rat/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bac_end/rat/bac_end_intro.html)  
Plate: 100 row: C column: 8  
Seq primer: T7  
Class: BAC ends.

## FEATURES

source

Location/Qualifiers  
1..442  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/strain="BN/SSNHsd/MCW"  
/db\_xref="taxon:10116"  
/clone="CH230-100C8"  
/sex="Female"  
/cell\_type="Brain"  
/clone\_id="CHORI-230 Segment 1"  
/note="Vector: pTARBAC2.1; Site\_1: EcoRI; Site\_2: EcoRI;  
CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by  
Pieter de Jong"

## ORIGIN

Query Match 38.1%; Score 114.6; DB 28; Length 442;  
Best Local Similarity 86.9%; Pred. No. 6.4e-11;  
Matches 126; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 1 TTTTCTTTTGTGAAGATTATTTTACTGCTGGAATCTTTGTAGAAACCAAG 60  
DB 382 TTATGCAATTAATTTCTCTTAAGACTATTTTACTGCGGAATACATTTTGAAGAAATC 323



QY 61 AGCAAGAACTTTTAAAGCTGGAGAGGCTGATGAAATTCAGTTCCTCCAGAC 120  
 DB 322 ATGAAAGAACTTTTAAAGCTGGAGAGGCTGATGAAATTCAGTTCCTCCAGAC 263  
 QY 121 AGCTTCGGCGCATCTTTTGTGTAAG 145  
 DB 262 AGCTTCGGCGCATCTTTTGTGTAAG 238

## RESULT 13

LOCUS BX402063 1201 bp mRNA linear EST 13-MAY-2003  
 DEFINITION BX402063 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED  
 Homo sapiens cDNA clone CS0D012YD18 5-PRIME, mRNA sequence.  
 ACCESSION BX402063  
 VERSION BX402063.1 GI:30626645  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1201)  
 Li, W.B., Gruber, C., Jesse, J. and Polayes, D.  
 Full-length cDNA libraries and normalization  
 Unpublished (2001)  
 COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 6672.r For  
 more information about this cluster, see  
 http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CS0D012YD1809091&cluster=6672.r. Contact :  
 Feng Liang Email : fliang@lifetech.com URL :  
 http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID : CS0D012YD1809091.

## FEATURES

## source

1..1201  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0D012YD18"  
 /cell\_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"  
 /cell\_line="RAMOS CELL LINE"  
 /clone\_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT  
 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo(dT)  
 primer. Five prime end enriched, double-strand cDNA was  
 digested with Not I and cloned into the Not I and EcoR V  
 sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

Query Match 35.6%; Score 107.2; DB 13; Length 1201;  
 Best Local Similarity 96.8%; Pred. No. 8e-10; 3; Indels 1; Gaps 1;  
 Matches 120; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 18 TTGAAGATTATTTTACTGCTGAATACT-TTGTAGAAAAACGAGAAAGAACTTCAA 76  
 DB 572 TTGAAGATTATTTTACTGCTGAATACTGTTGTAGAAAAACGAGAAAGAACTTCAA 631  
 QY 77 AGCTTCGGAGAGGCTGATGAAATTCAGTTCCTCCAGACAGCTTCGGCGCATCCT 136  
 DB 632 AGCTTCGGAGAGGCTGATGAAATTCAGTTCCTCCAGACAGCTTCGGCGCATCCT 691  
 QY 137 TTGG 140  
 DB 692 TTGG 695

## RESULT 14

BG758815

LOCUS BG758815 843 bp mRNA linear EST 15-MAY-2001  
 DEFINITION 602713177P1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4853467 5',  
 mRNA sequence.  
 ACCESSION BG758815  
 VERSION BG758815.1 GI:14069468  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 843)  
 NIH-MGC http://mgi.nci.nih.gov/.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: L1CM1699 row: 1 column: 20  
 High quality sequence stop: 615.

## FEATURES

## source

1..843  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4853467"  
 /tissue\_type="Primary B-cells from tonsils (cell line)"  
 /lab\_note="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_48"  
 /note="Organ: B-cells; Vector: pOTB7; Site 1: XhoI;  
 Site 2: EcoRI; cDNA made by oligo-dT priming.  
 Directionally cloned into EcoRI/XhoI sites using the  
 following 5' adaptor: GGGCAGAG(G). Size-selected >500bp  
 for average insert size 1.8kb. Library constructed by Ling  
 Hong in the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies).  
 Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 35.0%; Score 105.4; DB 12; Length 843;  
 Best Local Similarity 91.1%; Pred. No. 2e-09;  
 Matches 123; Conservative 0; Mismatches 11; Indels 1; Gaps 1;

QY 7 TTTTATTTTATTTTGAAGATTATTTTACTGCTGAATACTTTGTAGAAAAACGAGAA 66  
 DB 27 TTTTATTTTATTTTGAAGATTATTTTACTGCTGAATACTTTGTAGAAAAACGAGAA 86  
 QY 67 GAACCTT-CAAGCTCGGAGAGGCTGATGAAATTCAGTTCCTCCAGACAGCTT 125  
 DB 87 GAACCTTCAAGAGGCTGATGAAATTCAGTTCCTCCAGACAGCTT 146  
 QY 126 CGGCGCATCTTTTG 140  
 DB 147 CGGCGCATCTTTTG 161

## RESULT 15

BG144705

LOCUS BG144705 522 bp mRNA linear EST 01-FEB-2001  
 DEFINITION uc73f07.y1 Soares\_mouse\_mus musculus cDNA clone  
 IMAGE:333637 5' similar to TR:Q9NWE0 Q9NWE0 ACTIVATION-INDUCED  
 CYTIDINE DEAMINASE, mRNA sequence.  
 ACCESSION BG144705  
 VERSION BG144705.1 GI:12648105  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus



## REFERENCE

1 (bases 1 to 522)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index

## JOURNAL

Unpublished (1997)  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs@mail.nih.gov](mailto:cgapbs@mail.nih.gov)

This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
 MGI:1077801

Seq primer: -40RP from Gibco  
 High quality sequence stop: 487.

## FEATURES

## source

1..522  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:333637"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="Soares\_mouse\_MG8\_bcell"  
 /note="Organ: Germinal B-cell; Vector: p773D-Pac  
 (Pharmacia) with a modified polylinker; Site 1: Not I;  
 Site 2: Eco RI; 1st strand cDNA was primed with a Not I -  
 oligo(dT) primer [5'  
 TGTTACCAATCTGAGGTGGAGGCGCGCTGTTTCTTTTCTTTTCTTTTCTTTT  
 T 3']; double-stranded cDNA was ligated to Eco RI  
 adaptors (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of the modified p773 vector.  
 Library is normalized; constructed by Bento Soares and  
 M.Fatima Bonaldi."

## ORIGIN

## Query Match

33.4%; Score 100.6; DB 12; Length 522;

Best Local Similarity 88.6%; Pred.No.1.8e-08;  
 Matches 109; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY	18	TTTGAGATTATTTTCTGCTGGAATCTTTGTGAGAAACGAGAAAGACTTCAAA	77
DB	338	TTCAAAGACTATTTTCTGCTGGAATCAATTGTAGAAATCGTGAAGAACTTCAAA	397
QY	78	GCCTGGGAAGGCTGCAATGAAATTCAGTCCGCTCTCCAGACAGCTCGGCGCATCCTT	137
DB	398	GCCTGGGAAGGCTACATGAAATTCGTCGGCTAACGACACACTCGGCGCATCCTT	457
QY	138	TTG 140	
DB	458	TTG 460	

Search completed: March 13, 2004, 00:38:56  
 Job time : 962.585 secs



Result No.	Score	Query Match	Length	ID	Description
1	301	100.0	6564	9	US-09-966-8804-10
2	301	100.0	11204	9	US-09-966-8804-35
3	118.2	39.3	2818	9	US-09-966-8804-7
4	116	33.4	116	9	US-09-966-8804-14
5	100.6	38.5	2440	9	US-09-966-8804-1
6	47.8	15.9	475	10	US-09-918-995-31138
7	47.8	15.9	506	10	US-09-918-995-10105
8	47.6	15.8	344	12	US-10-424-593-3322
9	47.2	15.7	7657	14	US-10-339-676-185
10	47.2	15.7	7657	14	US-10-311-455-1995
11	47	15.6	337	9	US-09-960-352-6976
12	46.8	15.5	201143	12	US-10-840-425-1099
13	46.6	15.5	351	12	US-10-424-593-35643
14	46.6	15.5	476	12	US-10-424-593-28508
15	46.6	15.5	530	12	US-10-424-593-137880

C 16	46.6	15.5	2297	12	US-10-424-599-339282	Sequence 339282, A
C 17	46.4	15.4	2142	12	US-10-424-599-118533	Sequence 118533, A
C 18	46.4	15.4	2322	12	US-10-424-599-1231	Sequence 1231, Ap
C 19	46.2	15.4	531	12	US-10-424-599-84168	Sequence 84168, A
C 20	46.2	15.3	561	12	US-10-424-599-76196	Sequence 76196, A
C 21	46.2	15.3	601	12	US-10-424-599-59636	Sequence 59636, A
C 22	46	15.3	216	9	US-09-960-352-1902	Sequence 7902, Ap
C 23	46	15.3	399	10	US-09-918-599-17151	Sequence 17151, A
C 24	45.8	15.2	545	12	US-10-424-599-14829	Sequence 4829, A
C 25	45.8	15.2	6334	14	US-10-311-455-1185	Sequence 1185, Ap
C 26	45.6	15.1	391	12	US-10-424-599-48934	Sequence 48934, A
C 27	45.6	15.1	482	12	US-10-424-599-100328	Sequence 100328, A
C 28	45.6	15.1	660	12	US-10-424-599-24054	Sequence 24054, A
C 29	45.4	15.1	488	12	US-10-424-599-13762	Sequence 134762, A
C 30	45.4	15.1	5586	14	US-10-311-455-622	Sequence 622, Ap
C 31	45.4	15.1	37973	14	US-10-311-455-2159	Sequence 2159, Ap
C 32	45.2	15.0	655	12	US-10-424-599-107600	Sequence 107600, A
C 33	45.2	15.0	973	12	US-10-424-599-59879	Sequence 59879, A
C 34	45.2	15.0	1495	12	US-10-424-599-36589	Sequence 36589, A
C 35	45.2	15.0	1629	12	US-10-424-599-9786	Sequence 9786, Ap
C 36	45.2	15.0	1672	13	US-10-006-867-17	Sequence 17, Ap1
C 37	45.2	15.0	1672	13	US-10-006-867-17	Sequence 17, Ap1
C 38	45.2	15.0	1672	14	US-10-063-616-17	Sequence 17, Ap1
C 39	45.2	15.0	1672	14	US-10-063-502-17	Sequence 17, Ap1
C 40	45.2	15.0	1672	14	US-10-063-518-17	Sequence 17, Ap1
C 41	45.2	15.0	1672	14	US-10-063-598-17	Sequence 17, Ap1
C 42	45.2	15.0	1672	14	US-10-227-693-17	Sequence 17, Ap1
C 43	45.2	15.0	1672	14	US-10-063-567-17	Sequence 17, Ap1
C 44	45.2	15.0	1672	14	US-10-063-599-17	Sequence 17, Ap1
C 45	45.2	15.0	1672	14	US-10-063-599-17	Sequence 17, Ap1

## ALIGNMENTS

```

1 RESULT 1
2 US-09-966-880A-10
3 ; Sequence 10, Application US/09966880A
4 ; Patent No. US20020164743A1
5 ; GENERAL INFORMATION
6 ; APPLICANT: Honjo, Tasuku
7 ; APPLICANT: Muramatsu Masamichi
8 ; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
9 ; FILE REFERENCE: 06501-088001
10 ; CURRENT APPLICATION NUMBER: US/09/966,880A
11 ; CURRENT FILING DATE: 2001-09-28
12 ; PRIOR APPLICATION NUMBER: PCT/JP00/01918
13 ; PRIOR FILING DATE: 2000-03-28
14 ; PRIOR APPLICATION NUMBER: JP 11-771382
15 ; PRIOR FILING DATE: 1999-12-27
16 ; PRIOR APPLICATION NUMBER: JP 11-178959
17 ; PRIOR FILING DATE: 1999-06-24
18 ; PRIOR APPLICATION NUMBER: JP 11-87192
19 ; PRIOR FILING DATE: 1999-03-29
20 ; NUMBER OF SEQ ID NOS: 36
21 ; SOFTWARE: RastSeq for Windows Version 4.0
22 ; SEQ ID NO 10
23 ; LENGTH: 5654
24 ; TYPE: DNA
25 ; ORGANISM: Homo sapiens
26 US-09-966-880A-10

```

Query Match	100.0%	Score 301	DB 9	Length 6564
Best Local Similarity	100.0%	Pred. NO. 8.2e-6		
Matches 301	0	Mismatches 0	Indels 0	Gaps 0
Conservative				

Qy	Db
61	3132
ACGAAAGACTTTCAAGCCTGGGAGAGGGCTGCATGAAAAATTCACTTCGTCTCCAGAC	TTTTCTCTTTTTTTTTTTTTTTGAAGATTATTTTAACTGCTGGAAATCTTTGTAGAAAACC
120	3191
3192	
ACGGAAGAACTTTCAAAGCCTGGGAGAGGGCTGCATGAAAAATTCACTTCGTCTCCAGAC	
3251	



QY 121 ACCTTCGGGCGATCCTTTTGTAAAGGGGCTTCCTGCTTTTAAATTTCTTTCTC 180  
DB 3252 ACCTTCGGGCGATCCTTTTGTAAAGGGGCTTCCTGCTTTTAAATTTCTTTCTC 3311  
QY 181 TACAGCTTTTGTAGGTTTGTATATTTCTATATTTCTATATTTGTTCAATCACTTCA 240  
DB 3312 TACAGCTTTTGTAGGTTTGTATATTTCTATATTTCTATATTTGTTCAATCACTTCA 3371  
QY 241 GTTTCATCTGATGAAACCTTATTTCTCTCCATCAGCTTTTCTCTGCTGTTTCA 300  
DB 3372 GTTTCATCTGATGAAACCTTATTTCTCTCCATCAGCTTTTCTCTGCTGTTTCA 3431  
QY 301 C 301  
DB 3432 C 3432

## RESULT 2

US-09-966-880A-35  
Sequence 35, Application US/09966880A  
Patent No. US20020164743A1  
GENERAL INFORMATION:  
APPLICANT: Honjo, Tasuku  
APPLICANT: Muramatsu, Masamichi  
TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
FILE REFERENCE: 06501-088001  
CURRENT APPLICATION NUMBER: US/09/966,880A  
PRIOR FILING DATE: 2001-09-28  
PRIOR APPLICATION NUMBER: PCT/JP00/01918  
PRIOR FILING DATE: 2000-03-28  
PRIOR APPLICATION NUMBER: JP 11-371382  
PRIOR FILING DATE: 1999-12-27  
PRIOR APPLICATION NUMBER: JP 11-178999  
PRIOR FILING DATE: 1999-06-24  
PRIOR APPLICATION NUMBER: JP 11-87192  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 35  
LENGTH: 11204  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-966-880A-35

Query Match 100.0%; Score 301; DB 9; Length 11204;  
Best Local Similarity 100.0%; Pred. No. 1e-61;  
Matches 301; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTTTCCTTTTCTTTTCTTTTGAAGATATTTTACTGCTGGAATCTTTGAGAAAC 60  
DB 8347 TTTTCCTTTTCTTTTCTTTTGAAGATATTTTACTGCTGGAATCTTTGAGAAAC 8406  
QY 61 ACGAAAGACCTTCAAGGCTGGGAAGGCTGCATGAAATTCAGTCTCTCCAGAC 120  
DB 8407 ACGAAAGACCTTCAAGGCTGGGAAGGCTGCATGAAATTCAGTCTCTCCAGAC 8466  
QY 121 ACCTTCGGGCGATCCTTTTGTAAAGGGGCTTCCTGCTTTTAAATTTCTTTCTC 180  
DB 8467 ACCTTCGGGCGATCCTTTTGTAAAGGGGCTTCCTGCTTTTAAATTTCTTTCTC 8526  
QY 181 TACAGCTTTTGTAGGTTTGTATATTTCTATATTTCTATATTTGTTCAATCACTTCA 240  
DB 8527 TACAGCTTTTGTAGGTTTGTATATTTCTATATTTCTATATTTGTTCAATCACTTCA 8586  
QY 241 GTTTCATCTGATGAAACCTTATTTCTCTCCATCAGCTTTTCTCTGCTGTTTCA 300  
DB 8587 GTTTCATCTGATGAAACCTTATTTCTCTCCATCAGCTTTTCTCTGCTGTTTCA 8646  
QY 301 C 301  
DB 8647 C 8647

## RESULT 3

US-09-966-880A-7  
Sequence 7, Application US/09966880A  
Patent No. US20020164743A1  
GENERAL INFORMATION:  
APPLICANT: Honjo, Tasuku  
APPLICANT: Muramatsu, Masamichi  
TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
FILE REFERENCE: 06501-088001  
CURRENT APPLICATION NUMBER: US/09/966,880A  
PRIOR FILING DATE: 2001-09-28  
PRIOR APPLICATION NUMBER: PCT/JP00/01918  
PRIOR FILING DATE: 2000-03-28  
PRIOR APPLICATION NUMBER: JP 11-371382  
PRIOR FILING DATE: 1999-12-27  
PRIOR APPLICATION NUMBER: JP 11-178999  
PRIOR FILING DATE: 1999-06-24  
PRIOR APPLICATION NUMBER: JP 11-87192  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 7  
LENGTH: 2818  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (80)...(673)  
FEATURE:  
NAME/KEY: 5'UTR  
LOCATION: (1)...(79)  
FEATURE:  
NAME/KEY: 3'UTR  
LOCATION: (677)...(2818)  
US-09-966-880A-7

Query Match 39.3%; Score 118.2; DB 9; Length 2818;  
Best Local Similarity 97.6%; Pred. No. 2.5e-18;  
Matches 120; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 18 TTTGAAGATTTATTTTACTGCTGGAATCTTTTGTAGAAACCGAAGAACTTTCAA 77  
DB 500 TTTGAAGATTTATTTTACTGCTGGAATCTTTTGTAGAAACCGAAGAACTTTCAA 559  
QY 78 GCCTGGGAAGGCTGCATGAAATTCAGTCTCTCCAGACGCTTGGGCGATCCTT 137  
DB 560 GCCTGGGAAGGCTGCATGAAATTCAGTCTCTCCAGACGCTTGGGCGATCCTT 619

QY 138 TTG 140  
DB 620 TTG 622

## RESULT 4

US-09-966-880A-14  
Sequence 14, Application US/09966880A  
Patent No. US20020164743A1  
GENERAL INFORMATION:  
APPLICANT: Honjo, Tasuku  
APPLICANT: Muramatsu, Masamichi  
TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
FILE REFERENCE: 06501-088001  
CURRENT APPLICATION NUMBER: US/09/966,880A  
PRIOR FILING DATE: 2001-09-28  
PRIOR APPLICATION NUMBER: PCT/JP00/01918  
PRIOR FILING DATE: 2000-03-28  
PRIOR APPLICATION NUMBER: JP 11-371382  
PRIOR FILING DATE: 1999-12-27  
PRIOR APPLICATION NUMBER: JP 11-178999  
PRIOR FILING DATE: 1999-06-24  
PRIOR APPLICATION NUMBER: JP 11-87192  
PRIOR FILING DATE: 1999-03-29



NUMBER OF SEQ ID NOS: 36  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 14  
LENGTH: 116  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-966-880A-14

Query Match  
Best Local Similarity 100.0%; Score 116; DB 9; Length 116;  
Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 25 ATTATTTTACTGCTGGAATCTTTGTAGAAAACCAAGAACTTTCAAGCTGG 84  
DB 1 ATTATTTTACTGCTGGAATCTTTGTAGAAAACCAAGAACTTTCAAGCTGG 60  
QY 85 AAGGGCTGATGAAATTCAGTTCTCTCCAGACGCTTGGGGCATCCTTTTG 140  
DB 61 AAGGGCTGATGAAATTCAGTTCTCTCCAGACGCTTGGGGCATCCTTTTG 116

## RESULT 5

US-09-966-880A-1  
Sequence 1, Application US/09966880A  
Patent No. US20020164743A1  
GENERAL INFORMATION:  
APPLICANT: Honjo, Tasuku  
APPLICANT: Muramatsu, Masamichi  
TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
FILE REFERENCE: 06501-088001  
CURRENT APPLICATION NUMBER: US/09/966,880A  
PRIOR FILING DATE: 2001-09-28  
PRIOR APPLICATION NUMBER: PCT/JP00/01918  
PRIOR FILING DATE: 2000-03-28  
PRIOR APPLICATION NUMBER: JP 11-371382  
PRIOR FILING DATE: 1999-12-27  
PRIOR APPLICATION NUMBER: JP 11-178999  
PRIOR FILING DATE: 1998-06-24  
PRIOR APPLICATION NUMBER: JP 11-87192  
PRIOR FILING DATE: 1999-03-29  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1  
LENGTH: 2440  
TYPE: DNA  
ORGANISM: Mus musculus  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (93)...(686)  
FEATURE:  
NAME/KEY: 5'UTR  
LOCATION: (1)...(92)  
FEATURE:  
NAME/KEY: 3'UTR  
LOCATION: (690)...(2440)  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(2440)  
OTHER INFORMATION: n = A,T,C or G  
US-09-966-880A-1

Query Match  
Best Local Similarity 33.4%; Score 100.6; DB 9; Length 2440;  
Matches 109; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 18 TTGAAGATTATTTTACTGCTGGAATCTTTGTAGAAAACCAAGAACTTTCAA 77  
DB 513 TTCAAGAGACTATTTTACTGCTGGAATCTTTGTAGAAAACCAAGAACTTTCAA 572  
QY 78 GCTGGGAAAGGCTGATGAAATTCAGTTCTCTCCAGACGCTTGGGGCATCCTT 137  
DB 573 GCTGGGAAAGGCTGATGAAATTCAGTTCTCTCCAGACGCTTGGGGCATCCTT 632

QY 138 TTG 140  
DB 633 TTG 635

## RESULT 6

US-09-918-995-31138/c  
Sequence 31138, Application US/09918995  
Publication No. US20030073623A1  
GENERAL INFORMATION:  
APPLICANT: Hyseq, Inc.  
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
FROM VARIOUS CDNA LIBRARIES  
FILE REFERENCE: 20411-756  
CURRENT APPLICATION NUMBER: US/09/918,995  
PRIOR FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: US/09/235,076  
PRIOR FILING DATE: 1999-01-20  
NUMBER OF SEQ ID NOS: 38054  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 31138  
LENGTH: 475  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(475)  
OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-31138

Query Match  
Best Local Similarity 15.9%; Score 47.8; DB 10; Length 475;  
Matches 97; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 83 GGAAGGCTGATGAAATTCAGTTCTCTCCAGACGCTTGGGGCATCCTTTTGT 142  
DB 388 GGGGGGTTTAAAAAACCACCGAGGGGTTTCGGGGGGTTTTCOCCTTTT 329  
QY 143 AAGGGCTGCTGCTGCTTTTAAATTTCTCTCTCTCTCTCTCTCTCTCTCT 202  
DB 328 TTGGGGGCCCCNNNNNTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 269  
QY 203 TATATTTCTATATTTTCTATTTGTCAATCACTTCAGTTTCACTATGAACTT 262  
DB 268 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 209  
QY 263 ATT 265  
DB 208 TTT 206

## RESULT 7

US-09-918-995-10105/c  
Sequence 10105, Application US/09918995  
Publication No. US20030073623A1  
GENERAL INFORMATION:  
APPLICANT: Hyseq, Inc.  
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
FROM VARIOUS CDNA LIBRARIES  
FILE REFERENCE: 20411-756  
CURRENT APPLICATION NUMBER: US/09/918,995  
PRIOR FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: US/09/235,076  
PRIOR FILING DATE: 1999-01-20  
NUMBER OF SEQ ID NOS: 38054  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 10105  
LENGTH: 506  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(506)



OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-10105

Query Match 15.9%; Score 47.8; DB 10; Length 506;  
Best Local Similarity 58.6%; Pred. No. 0.084;  
Matches 82; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 127 GGGCGATCTTTGGTAAGGGGCTCCGCTTTTAAATTTCTTTCTTCTACAGT 186  
DB 324 GGGCTTATTTTAAAGGGGCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 265  
QY 187 CTTTGTGAGTTTGGATATTTCTTATATTTCTTATTTGTCATCAGTTTTC 246  
DB 264 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 205  
QY 247 ATCTGATGAAAATTATTT 266  
DB 204 GTTCCATATMAAACTTATTT 185

## RESULT 8

US-10-424-599-3232/c  
Sequence 3232, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 3232  
LENGTH: 344  
TYPE: DNA  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MBT3847\_102923C.1  
US-10-424-599-3232

Query Match 15.8%; Score 47.6; DB 12; Length 344;  
Best Local Similarity 54.6%; Pred. No. 0.081;  
Matches 95; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 127 GGGCGATCTTTGGTAAGGGGCTCCGCTTTTAAATTTCTTTCTTCTACAGT 186  
DB 198 GGGCGGCTCCAGAGGGGGGGGCGCCCTTTTCTTTTCTTTTCTTTTCTTTT 139  
QY 187 CTTTGTGAGTTTGGATATTTCTTATATTTCTTATTTGTCATCAGTTTTC 246  
DB 138 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 79  
QY 247 ATCTGATGAAAATTATTTCTTCTCCATCAGCTTTTCTTCTGCTGTTTC 300  
DB 78 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTCA 25

## RESULT 9

US-10-239-676-185  
Sequence 185, Application US/10239676  
Publication No. US20030082609A1  
GENERAL INFORMATION:  
APPLICANT: OLEK, Alexander  
APPLICANT: PIEPENBERCK, Christian  
APPLICANT: BERLIN, Kurt  
TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation  
FILE REFERENCE: 5013.1003  
CURRENT APPLICATION NUMBER: US/10/239,676  
CURRENT FILING DATE: 2002-09-24  
PRIOR APPLICATION NUMBER: PCT/EP01/03968

DE 10019058.8  
DE 10019173.8  
DE 10032529.7  
DE 10043826.1  
PRIOR FILING DATE: 2001-04-06  
2000-04-06  
2000-04-07  
2000-04-30  
2000-09-01  
2000-09-01

NUMBER OF SEQ ID NOS: 228  
SEQ ID NO 185

LENGTH: 7657  
TYPE: DNA

ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-239-676-185

Query Match 15.7%; Score 47.2; DB 14; Length 7657;  
Best Local Similarity 47.6%; Pred. No. 0.33;  
Matches 139; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

QY 7 TTTTCTTTTCTTTTGAAGATTATTTTACTGCTGGAATCTTTGTAGAAAACACGAAA 66  
DB 5513 TTGTATGATGATATATAGTATATGTTTATTTATATTTATGAGAGGATAT 5572  
QY 67 GAACCTTCAAGCTGGGAGGCTGCATGAAATTCAGTTGCTCCAGACAGCTTC 126  
DB 5573 TATTTTATGTAAGAGTGATGATGAAAGATGAAATTTATATATATTTT 5632  
QY 127 GGGCGATCTTTGGTAAGGGGCTCCGCTTTTAAATTTCTTTCTTCTACAGT 186  
DB 5633 TTTTATGATGATGATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 5692  
QY 187 CTTTGTGAGTTTGGATATTTCTTATATTTCTTATTTGTCATCAGTTTTC 246  
DB 5693 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 5752  
QY 247 ATCTGATGAAAATTATTTCTTCTCCATCAGCTTTTCTTCTGCTGTTTC 298  
DB 5753 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 5804

## RESULT 10

US-10-311-455-1995  
Sequence 1995, Application US/10311455  
Publication No. US20030143606A1  
GENERAL INFORMATION:  
APPLICANT: OLEK, Alexander  
APPLICANT: PIEPENBERCK, Christian  
APPLICANT: BERLIN, Kurt  
TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining  
TITLE OF INVENTION: Cytosine methylation  
FILE REFERENCE: 5013.1014  
CURRENT APPLICATION NUMBER: US/10/311,455  
CURRENT FILING DATE: 2002-12-16  
PRIOR APPLICATION NUMBER: PCT/EP01/07537  
PRIOR FILING DATE: 2001-07-02  
PRIOR APPLICATION NUMBER: DE 10032529.7  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: DE 10043826.1  
PRIOR FILING DATE: 2000-09-01  
NUMBER OF SEQ ID NOS: 2424  
SEQ ID NO 1995  
LENGTH: 7657  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-311-455-1995

Query Match 15.7%; Score 47.2; DB 14; Length 7657;  
Best Local Similarity 47.6%; Pred. No. 0.33;























STATE: MO  
COUNTRY: USA  
ZIP: 63105  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/897,126  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/520,678  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Henderson, Melodie W.  
REGISTRATION NUMBER: 37,848  
REFERENCE/DOCKET NUMBER: 6029-6836  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 314-727-5188  
TELEFAX: 314-727-6092  
TELEX:  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 270 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-897-126-30  
Query Match  
Best Local Similarity 13.7%; Score 41.2; DB 3; Length 270;  
Matches 73; Conservative 0; Mismatches 53; Indels 0; Gaps 0;  
QY 142 TAAGGGCTTCCTGCTTTAAATTTCTTCTCTACAGTCTTTTGAGATTTC 201  
DB 107 TTAGGCTCTCTGTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 166  
QY 202 GTATATTTCTTAATTTCTTAATGTTCAATCAGTCTGATTTGATGAAACTT 261  
DB 167 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 226  
QY 262 TATTTT 267  
DB 227 TTTTTC 232  
RESULT 8  
US-09-601-537-10/C  
Sequence 10, Application US/09601537  
Patent No. 6582691  
GENERAL INFORMATION:  
APPLICANT: Gallert, Karl-Christian  
APPLICANT: Mullner, Stefan  
APPLICANT: Huls, Christoph  
APPLICANT: Bohmsch, Britta  
TITLE OF INVENTION: Expression Vector for the Production of  
FILE REFERENCE: 50186/002001  
CURRENT APPLICATION NUMBER: US/09/601,537  
CURRENT FILING DATE: 2000-08-02  
PRIOR APPLICATION NUMBER: PCT/EP99/00829  
PRIOR FILING DATE: 1999-02-09  
PRIOR APPLICATION NUMBER: DE 198 05 781.4  
PRIOR FILING DATE: 1998-02-12  
NUMBER OF SEQ ID NOS: 29  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 10  
LENGTH: 441  
TYPE: DNA  
ORGANISM: Artificial Sequence

FEATURE:  
OTHER INFORMATION: p153-NT3  
US-09-601-537-10  
Query Match  
Best Local Similarity 13.6%; Score 40.8; DB 4; Length 441;  
Matches 75; Conservative 0; Mismatches 57; Indels 0; Gaps 0;  
QY 158 TTTTAAATTTCTTCTTCTCTACAGTCTTTTGGAGTTGATATTTCTATATT 217  
DB 425 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 366  
QY 218 TTTCTATGTTCAATCAGTCTGATTTGATGAAACTTATTTCTCTCCACT 277  
DB 365 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 306  
QY 278 CAGCTTTTCTT 289  
DB 305 AATTAATTTATT 294  
RESULT 9  
US-09-800-729-78/C  
Sequence 78, Application US/09800729  
Patent No. 6605592  
GENERAL INFORMATION:  
APPLICANT: N1 et al.  
TITLE OF INVENTION: 32 Human secreted proteins  
FILE REFERENCE: P2044P1  
CURRENT APPLICATION NUMBER: US/09/800,729  
CURRENT FILING DATE: 2001-03-08  
PRIOR APPLICATION NUMBER: PCT/US00/26013  
PRIOR FILING DATE: 2000-09-22  
PRIOR APPLICATION NUMBER: 60/155,709  
PRIOR FILING DATE: 1999-09-24  
NUMBER OF SEQ ID NOS: 217  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 78  
LENGTH: 1141  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-800-729-78  
Query Match  
Best Local Similarity 13.6%; Score 40.8; DB 4; Length 1141;  
Matches 63; Conservative 0; Mismatches 47; Indels 0; Gaps 0;  
QY 158 TTTTAAATTTCTTCTTCTCTACAGTCTTTTGGAGTTGATATTTCTATATT 217  
DB 1131 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1072  
QY 218 TTTCTATGTTCAATCAGTCTGATTTGATGAAACTTATTTCTCTCTCC 273  
DB 1071 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1016  
RESULT 10  
US-09-601-537-9/C  
Sequence 9, Application US/09601537  
Patent No. 6582691  
GENERAL INFORMATION:  
APPLICANT: Gallert, Karl-Christian  
APPLICANT: Mullner, Stefan  
APPLICANT: Huls, Christoph  
APPLICANT: Bohmsch, Britta  
TITLE OF INVENTION: Expression Vector for the Production of  
FILE REFERENCE: 50186/002001  
CURRENT APPLICATION NUMBER: US/09/601,537  
CURRENT FILING DATE: 2000-08-02  
PRIOR APPLICATION NUMBER: PCT/EP99/00829  
PRIOR FILING DATE: 1999-02-09  
PRIOR APPLICATION NUMBER: DE 198 05 781.4







RESULT 13  
US-09-621-976-1322/c  
Sequence 1322, Application US/09621976  
Patent No. 6639063  
GENERAL INFORMATION:  
APPLICANT: Dumas Milne Edwards, J.B.  
APPLICANT: Jobert, S.  
APPLICANT: Giordano, J.Y.  
TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
FILE REFERENCE: GENSET 054PR2  
CURRENT APPLICATION NUMBER: US/09/621,976  
NUMBER OF SEQ ID NOS: 19335  
SOFTWARE: Patent.pm  
SEQ ID NO 1322  
LENGTH: 249  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 28..249  
US-09-621-976-1322

Query Match 13.3%; Score 40; DB 4; Length 249;  
Best Local Similarity 57.0%; Pred. No. 0.14;  
Matches 73; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 158 TTTTAAATTTCTTCTTCTCTACAGCTTTTGGAGTTTGGTAATTTCTAATTT 217  
DB 247 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 188  
QY 218 TTTCTTTTCTTCAATCCTCAGTTTTCATCTGATGAAATTTATTTCTCCACAT 277  
DB 187 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 128  
QY 278 CAGCTTTT 285  
DB 127 CAATTTT 120

RESULT 14  
US-09-014-969-14/c  
Sequence 14, Application US/09014969  
Patent No. 5965397  
GENERAL INFORMATION:  
APPLICANT: Jacobs, Kenneth  
APPLICANT: McCoy, John M.  
APPLICANT: Lavallie, Edward R.  
APPLICANT: Racine, Lisa A.  
APPLICANT: Metberg, David  
APPLICANT: Treacy, Maurice  
APPLICANT: Spaulding, Vikki  
APPLICANT: Agostino, Michael J.  
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
TITLE OF INVENTION: ENCODING THEM  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 CambridgePark Drive  
CITY: Cambridge  
STATE: MA  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/014,969  
FILING DATE:

CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Sprunger, Suzanne A.  
REGISTRATION NUMBER: 41,323  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 498-8284  
TELEFAX: (617) 876-5851  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2447 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-09-014-969-14

Query Match 13.3%; Score 40; DB 2; Length 2447;  
Best Local Similarity 59.8%; Pred. No. 0.27;  
Matches 67; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 158 TTTTAAATTTCTTCTTCTCTACAGCTTTTGGAGTTTGGTAATTTCTAATTT 217  
DB 2335 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 2276  
QY 218 TTTCTTTTCTTCAATCCTCAGTTTTCATCTGATGAAATTTATTTCTC 269  
DB 2275 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 2224

RESULT 15  
US-10-204-708-23  
Sequence 23, Application US/10204708  
Patent No. 6677731  
GENERAL INFORMATION:  
APPLICANT: OLEK, Alexander  
APPLICANT: PLEBENROCK, Christian  
APPLICANT: BERLIN, Kurt  
TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication  
FILE REFERENCE: 5013.1012  
CURRENT APPLICATION NUMBER: US/10/204,708  
PRIOR FILING DATE: 2003-05-06  
PRIOR APPLICATION NUMBER: PCT/EP01/03971  
PRIOR FILING DATE: 2001-04-06  
PRIOR APPLICATION NUMBER: DE 10019058.8  
PRIOR FILING DATE: 2000-04-06  
PRIOR APPLICATION NUMBER: DE 10019173.8  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: DE 10032529.7  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: DE 10043826.1  
PRIOR FILING DATE: 2000-09-01  
NUMBER OF SEQ ID NOS: 98  
SEQ ID NO 23  
LENGTH: 11049  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-204-708-23

Query Match 13.3%; Score 40; DB 4; Length 11049;  
Best Local Similarity 53.9%; Pred. No. 0.42; 70; Indels 0; Gaps 0;  
Matches 82; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 101 TTTCAATTCCTCTCCAGACAGCTTGGGCAATCTTTTGGTAAGGAGCTTCTGCTTT 160  
DB 7020 TTTAATTTTATTTTAGAAGATTGGTTTTTGGTATTTAGATATTAAGTGAATTT 7079  
QY 161 TTTAATTTCTTCTTCTCTCAAGCTTTTGGAGTTTGGTAATTTCTAATTTTC 220  
DB 7080 TTTGAATTTTCTTTTATTTTATTTAGTTTGTGTTTGTGTTTGTGTTTGTGTTT 7139



Qy 221 TTATGTGCAATCACTGAGTTTCACTGA 252  
Db 7140 TTTAAATTATTTAGTTGAGTTTGAAGTGA 7171

Search completed: March 13, 2004, 00:51:04  
Job time : 25.3958 secs







CC The present invention describes an activation-induced cytidine deaminase  
CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and has  
CC cytidine activity similar to APOBEC-1. AID has anti-allergic, anti-nausea,  
CC antiasthmatic, ophthalmological, anti-HIV and dermatological activities,  
CC and can be used in gene therapy. AID polynucleotides are useful in  
CC methods for identifying drugs for the treatment of B cell associated  
CC immunodeficiency disorders, immunodeficiency diseases and allergies, such as  
CC immunoglobulin A (IgA) deficiency disease, IgA nephritis, gamma-  
CC globulinemia, atopic dermatitis, allergic colitis, asthma, food allergy,  
CC drug allergy, allergic rhinitis, Rosen disease, DiGeorge disease, ataxia  
CC telangiectasia, common variable immunodeficiency disorder, MHC (major  
CC histocompatibility class II deficiency disease, AIDS (auto  
CC immunodeficiency syndrome), elevated IgE disorder, and IgG subclass  
CC selection disorder. The DNA sequences encoding AID may be used for gene  
CC therapy and the antibodies to the AID protein may be used for diagnosis  
CC and treatment of these disorders. The present sequence represents a  
CC genomic DNA sequence of human AID  
CC  
XX

Sequence 6564 BP; 1909 A; 1358 C; 1383 G; 1914 T; 0 U; 0 Other;

Query Match 100.0%; Score 301; DB 3; Length 6564;  
Best Local Similarity 100.0%; Pred. No. 5,3e-56;  
Matches 301; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTTTCCTTTTGTGAGATTATTTTCTGCTGGAATCTTTGTAGAAACC 60  
DB 3132 TTTTCCTTTTGTGAGATTATTTTCTGCTGGAATCTTTGTAGAAACC 3191  
QY 61 ACGAAGAACTTCAAGCCTGGAGGGGCTGCATGAAATTCAGTTCCTCTCCAGAC 120  
DB 3192 ACGAAGAACTTCAAGCCTGGAGGGGCTGCATGAAATTCAGTTCCTCTCCAGAC 3251  
QY 121 AGCTTCGGCGCATCCTTTGTAAGGGGCTCTCGCTTTTAAATTTCTTTCTTC 180  
DB 3252 AGCTTCGGCGCATCCTTTGTAAGGGGCTCTCGCTTTTAAATTTCTTTCTTC 3311  
QY 181 TACAGCTTTTGTAGAGTTTGTATTTCTTATTTCTTATTTCTTATTCATCAGCTCA 240  
DB 3312 TACAGCTTTTGTAGAGTTTGTATTTCTTATTTCTTATTTCTTATTCATCAGCTCA 3371  
QY 241 GTTTCATCTGATGAAACTTTATTTCTCTCCACATCAGCTTTTCTTCTGCTTTCA 300  
DB 3372 GTTTCATCTGATGAAACTTTATTTCTCTCTCCACATCAGCTTTTCTTCTGCTTTCA 3431  
QY 301 C 301  
DB 3432 C 3432

RESULT 2  
AAC55339  
ID AAC55339 standard; DNA; 11204 BP.

XX AAC55339;

DT 05-FEB-2001 (first entry)

DE Human activation-induced cytidine deaminase genomic DNA SEQ ID NO:35.

XX Activation-induced cytidine deaminase; AID; cytidine deaminase;  
XX Immune related disease; allergy; allergic disease; anti-allergic;  
XX antiasthmatic; antiasthmatic; ophthalmological; anti-HIV; dermatological;  
XX gene therapy; B cell associated immune system disorder; food allergy;  
XX immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
XX IgA nephritis; gamma-globulinemia; atopic dermatitis; allergic colitis;  
XX drug allergy; allergic rhinitis; Rosen disease; DiGeorge disease; AIDS;  
XX ataxia telangiectasia; common variable immunodeficiency disorder;  
XX major histocompatibility class II deficiency disease;  
XX auto immunodeficiency syndrome; IgG subclass selection disorder; ds.

OS Homo sapiens.

XX WO200058480-A1.

XX 05-OCT-2000.  
PD 28-MAR-2000; 2000MO-VP001918.  
XX  
XX 29-MAR-1999; 99JP-00087192.  
PR 24-JUN-1999; 99JP-00178999.  
PR 27-DEC-1999; 99JP-00371382.  
XX  
XX (NIBS) JAPAN TOBACCO INC.  
PA (HONJ) HONJO T.  
XX  
XX Honjo T, Muramatsu M;  
XX WPI; 2000-611715/58.  
XX  
XX Nucleic acid encoding activation induced cytidine deaminase, useful as a  
XX target for drug development for immune-related diseases including  
XX allergies.  
XX  
XX Claim 17; Page 163-170; 174pp; Japanese.

CC The present invention describes an activation-induced cytidine deaminase  
CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and has  
CC cytidine activity similar to APOBEC-1. AID has anti-allergic, anti-nausea,  
CC antiasthmatic, ophthalmological, anti-HIV and dermatological activities,  
CC and can be used in gene therapy. AID polynucleotides are useful in  
CC methods for identifying drugs for the treatment of B cell associated  
CC immunodeficiency disorders, immunodeficiency diseases and allergies, such as  
CC immunoglobulin A (IgA) deficiency disease, IgA nephritis, gamma-  
CC globulinemia, atopic dermatitis, allergic colitis, asthma, food allergy,  
CC drug allergy, allergic rhinitis, Rosen disease, DiGeorge disease, ataxia  
CC telangiectasia, common variable immunodeficiency disorder, MHC (major  
CC histocompatibility class II deficiency disease, AIDS (auto  
CC immunodeficiency syndrome), elevated IgE disorder, and IgG subclass  
CC selection disorder. The DNA sequences encoding AID may be used for gene  
CC therapy and the antibodies to the AID protein may be used for diagnosis  
CC and treatment of these disorders. The present sequence represents a  
CC genomic DNA sequence of human AID  
CC  
XX

Sequence 11204 BP; 3305 A; 2273 C; 2373 G; 3253 T; 0 U; 0 Other;

Query Match 100.0%; Score 301; DB 3; Length 11204;  
Best Local Similarity 100.0%; Pred. No. 5,6e-56;  
Matches 301; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTTTCCTTTTGTGAGATTATTTTCTGCTGGAATCTTTGTAGAAACC 60  
DB 8347 TTTTCCTTTTGTGAGATTATTTTCTGCTGGAATCTTTGTAGAAACC 8406  
QY 61 ACGAAGAACTTCAAGCCTGGAGGGGCTGCATGAAATTCAGTTCCTCTCCAGAC 120  
DB 8407 ACGAAGAACTTCAAGCCTGGAGGGGCTGCATGAAATTCAGTTCCTCTCCAGAC 8466  
QY 121 AGCTTCGGCGCATCCTTTGTAAGGGGCTCTCGCTTTTAAATTTCTTTCTTC 180  
DB 8467 AGCTTCGGCGCATCCTTTGTAAGGGGCTCTCGCTTTTAAATTTCTTTCTTC 8526  
QY 181 TACAGCTTTTGTAGAGTTTGTATTTCTTATTTCTTATTTCTTATTTCTTATTCATCAGCTCA 240  
DB 8527 TACAGCTTTTGTAGAGTTTGTATTTCTTATTTCTTATTTCTTATTTCTTATTCATCAGCTCA 8586  
QY 241 GTTTCATCTGATGAAACTTTATTTCTCTCCACATCAGCTTTTCTTCTGCTTTCA 300  
DB 8587 GTTTCATCTGATGAAACTTTATTTCTCTCCACATCAGCTTTTCTTCTGCTTTCA 8646  
QY 301 C 301  
DB 8647 C 8647

RESULT 3  
ABS73286



```

ID  ABS73286 standard; DNA; 11204 BP
XX
XX  ABS73286;
XX
XX  04-DEC-2002 (first entry)
XX
XX  DNA encoding human translocation del(12p) protein #1.
DE
XX
XX  Chromosome aberration; oncogenic fusion protein; cancer;
XX  proliferative disease; cellular protein isoform; heat shock protein 90;
XX  HSP-90; rheumatoid arthritis; cancer; haematopoietic disorder;
XX  T cell lymphoma; B cell lymphoma; chronic myeloid leukaemia; CML;
XX  acute myeloid leukaemia; AML; chronic myelomonocytic leukaemia; CMML;
XX  acute lymphoblastic leukaemia; ALL; APL; NHL; solid tumour;
XX  papillary thyroid carcinoma; Ewing's sarcoma; melanoma; liposarcoma;
XX  rhabdomyosarcoma; synovial sarcoma; viral infection; gene; ds.
XX
XX  Homo sapiens.
OS
XX
XX  WO200269900-A2.
XX
XX  12-SEP-2002.
XX
XX  01-MAR-2002; 2002WO-US006518.
XX
XX  01-MAR-2001; 2001US-0272751P.
XX
XX  (CONF-) CONFORMA THERAPEUTICS CORP.
XX
XX  Filtz LC, Burrows FU;
XX
XX  WPI: 2002-698710/75.
XX
XX  P-PSDB; ABG95082.
XX
XX  Treating genetically-defined disease associated with chromosomal
XX  aberrations yielding oncogenic fusion proteins, e.g. cell proliferative
XX  diseases, involves administering an inhibitor of heat shock protein 90.
XX
XX  Disclosure; Page 242-245; 389pp; English.
XX
XX  The invention describes a method of treating genetically-defined disease
XX  associated with chromosomal aberrations yielding oncogenic fusion
XX  proteins (I), treating cancerous cells containing (I) in a heterogeneous
XX  cell population, treating proliferative diseases associated with mutant
XX  protein or cellular protein isoforms (II) dependent on heat shock protein
XX  (HSP)-90, or selectively treating cells expressing (II) involving
XX  administering HSP90-inhibitor. The method is useful for treating
XX  genetically-defined disease with chromosomal aberration yielding
XX  oncogenic fusion protein, treating cancerous cells containing fusion
XX  protein in heterogeneous cell population, treating proliferative disease
XX  (e.g. rheumatoid arthritis or cancer) associated with mutant protein or
XX  cellular protein isoform dependent on heat shock protein (HSP)-90 (e.g.
XX  p53), or selectively treating cells expressing mutant protein or cellular
XX  protein isoform in a patient heterozygous for (II). The method is useful
XX  for treating a disease e.g. haematopoietic disorder such as T or B cell
XX  lymphoma, chronic myeloid leukaemia (CML), APL, ALL, NHL and CMML,
XX  or a disease characterised by a solid tumour such as papillary thyroid
XX  carcinoma, Ewing's sarcoma, melanoma, liposarcoma, rhabdomyosarcoma and
XX  synovial sarcoma. The method is also useful for treating viral
XX  infections. This represents the DNA sequence of a chromosome aberration
XX
XX  Sequence 11204 BP; 3305 A; 2273 C; 2373 G; 3253 T; 0 U; 0 Other;
SQ
XX
XX  Query Match 100.0%; Score 301; DB 6; Length 11204;
XX  Best Local Similarity 100.0%; Pred. NO. 5.6e-56;
XX  Matches 301; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX  1 TTTTCCTTTTCTTTTGTGAAGATATTTTACTGCGGAAATCTTTGTAGAAAAC 60
XX  Db 8347 TTTTCCTTTTCTTTTGTGAAGATATTTTACTGCGGAAATCTTTGTAGAAAAC 8406
XX  61 ACGAAGAACTTCAAGCTTGGGAAAGGCTGCAATGAAATTCAGTTGCTCTTCAGAC 120

```

ID	Accession	Sequence	Length
Db	ACGAAGAAGCTTTAAAGCCGGGAAGGGCTGCATGAAATTCAGTTCGTCTCCACAGC	8467	
Qy	121 AGCTTCGGCCATCCCTTTGGTAGAGGGCTTCCTCGCTTTAAATTTCTTTCTTC	180	
Db	8467 AGCTTCGGCCATCCCTTTGGTAGAGGGCTTCCTCGCTTTAAATTTCTTTCTTC	8526	
Qy	181 TACAGCTTTTGGAGTTGGATTCCTATATTTCTTATTTGTTCAATCACTCA	240	
Db	8527 TACAGCTTTTGGAGTTGGATTCCTATATTTCTTATTTGTTCAATCACTCA	8586	
Qy	241 GTTTTCATCTGATGAAACCTTATTTCTTCCTCCACATCAGCTTTTCTTCGTGTTCA	300	
Db	8587 GTTTTCATCTGATGAAACCTTATTTCTTCCTCCACATCAGCTTTTCTTCGTGTTCA	8646	
Qy	301 C 301		
Db	8647 C 8647		
RESULT 4			
ID	ADB88952	standard; DNA; 597 BP.	
XX	ADB88952		
XX	ADB88952;		
AC			
DT	04-DEC-2003	(first entry)	
XX			
DE	AID gene as substrate used in AID mutagenic method.		
XX			
KW	ds; cytosstatic; virucide; antidote; mutation;		
KW	activation-induced cytidine deaminase; cell phenotype;		
KW	monoclonal antibody; antigen; cross-reactivity; tumour;		
XX	germ warfare agent; toxin; V region.		
XX			
OS	Homo sapiens.		
XX			
FX	Key	Location/Qualifiers	
FT	mutation	replace(9,A)	
FT		/*tag= a	
FT		/note= "mutation in PI-5 hybridoma cell"	
FT	mutation	replace(12,A)	
FT		/*tag= b	
FT		/note= "mutation in Ramos cell"	
FT	mutation	replace(14,C)	
FT		/*tag= c	
FT		/note= "mutation in CHO cell"	
FT	mutation	replace(61,T)	
FT		/*tag= d	
FT		/note= "mutation in PI-5 hybridoma cell"	
FT	mutation	replace(66,A)	
FT		/*tag= e	
FT		/note= "mutation in PI-5 hybridoma cell"	
FT	mutation	replace(73,T)	
FT		/*tag= f	
FT		/note= "mutation in CHO cell"	
FT	mutation	replace(79,G)	
FT		/*tag= g	
FT		/note= "mutation in Ramos cell"	
FT	mutation	replace(85,G)	
FT		/*tag= h	
FT		/note= "mutation in PI-5 hybridoma cell"	
FT	mutation	replace(112,C)	
FT		/*tag= i	
FT		/note= "mutation in Ramos cell"	
FT	mutation	replace(118,G)	
FT		/*tag= j	
FT		/note= "mutation in PI-5 hybridoma cell"	
FT	mutation	replace(156,T)	
FT		/*tag= k	
FT		/note= "mutation in PI-5 hybridoma cell"	
FT	mutation	replace(161,A)	
FT		/*tag= l	
FT		/note= "mutation in PI-5 hybridoma cell"	



```

FT mutation replace(165,T,G)
FT /*tag= m
FT /note= "G mutation in Ramos cell, T mutation in CHO cell"
FT replace(178,T)
FT /*tag= n
FT /note= "mutation in Ramos cell"
FT replace(206,G)
FT /*tag= o
FT /note= "mutation in CHO cell"
FT replace(209,C)
FT /*tag= p
FT /note= "mutation in Ramos cell"
FT replace(218,T)
FT /*tag= q
FT /note= "mutation in Ramos cell"
FT replace(250,C)
FT /*tag= r
FT /note= "mutation in Ramos cell"
FT replace(261,G)
FT /*tag= s
FT /note= "mutation in Ramos cell"
FT replace(289,A)
FT /*tag= t
FT /note= "mutation in Ramos cell"
FT replace(303,T)
FT /*tag= u
FT /note= "mutation in Ramos cell"
FT replace(329,T)
FT /*tag= v
FT /note= "mutation in Ramos cell"
FT replace(350,G,T)
FT /*tag= w
FT /note= "T mutation in Ramos cell, G mutation in CHO cell"
FT replace(356,A)
FT /*tag= x
FT /note= "A mutation in Ramos cell, A mutation in CHO cell"
FT replace(357,G)
FT /*tag= y
FT /note= "mutation in Ramos cell"
FT replace(367,T, A)
FT /*tag= z
FT /note= "mutation in PI-5 hybridoma cell"
FT replace(375,A)
FT /*tag= aa
FT /note= "mutation in CHO cell"
FT replace(402,A)
FT /*tag= ab
FT /note= "mutation in PI-5 hybridoma cell"
FT replace(410,T)
FT /*tag= ac
FT /note= "mutation in PI-5 hybridoma cell"
FT replace(418,T)
FT /*tag= ad
FT /note= "mutation in PI-5 hybridoma cell"
FT replace(420,T)
FT /*tag= ae
FT /note= "mutation in PI-5 hybridoma cell"
FT replace(421,A)
FT /*tag= af
FT /note= "mutation in CHO cell"
FT replace(430,G)
FT /*tag= ag
FT /note= "mutation in Ramos cell"
FT replace(440,A)
FT /*tag= ah
FT /note= "mutation in CHO cell"
FT replace(449,T)
FT /*tag= ai
FT /note= "mutation in PI-5 hybridoma cell"
FT replace(495,A)
FT /*tag= aj
FT /note= "mutation in CHO cell"
FT replace(517,C)

```

```

FT /*tag= ak
FT /note= "mutation in Ramos cell"
FT replace(526,T, A)
FT /*tag= al
FT /note= "T mutation in PI-5 hybridoma cell, A mutation in
FT Ramos cell"
FT replace(533,A)
FT /*tag= am
FT /note= "mutation in PI-5 hybridoma cell"
FT replace(538,A)
FT /*tag= an
FT /note= "mutation in CHO cell"
FT replace(549,A)
FT /*tag= ao
FT /note= "mutation in CHO cell"
FT replace(568,T)
FT /*tag= ap
FT /note= "mutation in Ramos cell"
FT replace(570,T)
FT /*tag= aq
FT /note= "mutation in CHO cell"

MO2003061363-A2.
31-JUL-2003.
15-JAN-2003; 2003MO-US001149.
17-JAN-2002; 2002US-0350269P.
(YASH ) UNIV YESHIYA EINSTEIN COLLEGE.
Martin A, Scharif MD;
WPI; 2003-747983/70.
Induction of mutations in gene expressed in eukaryotic cell useful e.g.
to produce antibodies with greater affinity or specificity for antigens
e.g. human pathogens by expressing an activation-induced cytidine
deaminase gene in cell.
Example 4; Fig 6; 73pp; English.
The invention relates to the induction of mutations in a gene expressed
in a eukaryotic cell, where the gene is operably linked to a promoter and
within 2 kb of promoter, by expressing a transgenic activation-induced
cytidine deaminase (Aid) gene in the cell. The method is useful for
producing mutated genes and proteins, determining the effect of a
mutation on a protein or cell phenotype, and producing genes and proteins
with altered properties. It is especially useful for producing antibodies
CC with altered affinities or specificities for an antigen, or
CC increased/decreased cross-reactivity for a second antigen, e.g. to
CC produce antibodies with greater affinity/specificity useful
CC therapeutically to treat tumours, viruses such as Ebola and Lassa Fever
CC or against germ warfare agents, toxins (e.g. ricin). The method may
CC similarly be used to alter affinity, specificity or cross-reactivity of a
CC monoclonal antibody. This sequence represents the Aid gene used in the
CC method of the invention to induce mutations in the sequence.
XX
XX Sequence 597 BP; 128 A; 164 C; 155 G; 150 T; 0 U; 0 Other;
SQ
Query Match 39.3%; Score 118.2; DB 8; Length 597;
Best Local Similarity 97.6%; Pred. No. 1.8e-16;
Matches 120; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 18 TTGGAAGTATTTTACTGCTGGAATPCTTTGTAGAAAACCGAAGAACTTTGAA 77
DB 421 TTCAAGATATTTTACTGCTGGAATPCTTTGTAGAAAACCGAAGAACTTTGAA 480
QY 78 GCGTGGGAAGGCGCGCAAGAAATTCAGTGTCTCCAGACAGCTTCGCGCATCCTT 137
DB 481 GCGTGGGAAGGCGCGCAAGAAATTCAGTGTCTTCAGACAGCTTCGCGCATCCTT 540

```



QY 138 TTG 140  
DB 541 TTG 543

RESULT 5  
ABX05468  
ID ABX05468 standard; cDNA; 1543 BP.  
XX  
XX ABX05468;  
XX  
XX 17-JAN-2003 (first entry)  
XX  
XX Human novel polynucleotide #483.  
XX  
XX Human; gene; ss; genetic disorder; gene mapping; medical imaging; cancer;  
KM neurodegenerative disorder; lymphoid cell disorder; osteoporosis;  
KM Parkinson's disease; Alzheimer's disease; bone degenerative disorder;  
KM osteoarthritis; periodontal disease; liver fibrosis; viral infection;  
KM fungal infection; bacterial infection; autoimmune disease; diabetes;  
KM atopic dermatitis.  
XX  
XX Homo sapiens.  
XX  
XX WO200274961-A1.  
XX  
XX 26-SEP-2002.  
XX  
XX 14-MAR-2002; 2002WO-US005109.  
XX  
XX 15-MAR-2001; 2001US-00810173.  
XX  
XX (HYSEQ-) HYSEQ INC.  
XX  
XX Tang YT, Zhou P, Goodrich R, Asundi V, Zhang J, Zhao QA, Ren F,  
PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M,  
PI Wehrman T, Wang J, Wang D, Drmanac RT;  
XX  
XX WPI: 2003-040556/03.  
XX  
XX P-PSDB; AB000390.  
XX  
XX New isolated polypeptides and polynucleotides, useful for preventing,  
PT treating or ameliorating medical conditions, such as cancer,  
PT neurodegenerative disorders, lymphoid cell disorders, bone degenerative  
PT disorders, and infections.  
XX  
XX Claim 1; SEQ ID NO 483; 235pp; English.  
XX  
XX The invention relates to human polynucleotides and the polypeptides they  
XX encode. The polynucleotides and polypeptides are useful in diagnostics,  
XX forensics, gene mapping, medical imaging, identification of mutations,  
XX responsible for genetic disorders or other traits, assessing biodiversity  
XX and producing many other types of data and products dependent on DNA and  
XX amino acid sequences. They are also useful for preventing, treating or  
XX ameliorating medical conditions, such as cancer, neurodegenerative  
XX disorders (e.g. Parkinson's disease, Alzheimer's disease), lymphoid cell  
XX disorders, osteoporosis, osteoarthritis, bone degenerative disorders,  
XX periodontal disease, liver fibrosis, infections (e.g. viral, fungal or  
XX bacterial) or autoimmune diseases (e.g. diabetes, atopic dermatitis).  
XX Sequences ABX04986-ABX05511 represent human polynucleotides of the  
XX invention. Note: The sequence data for this patent is not represented in  
XX the printed specification but is based on sequence information supplied  
XX by the European Patent Office

QY 18 TTGAAGATTATTTTTCCTGCTGGAATACCTTTGTAGAAAACGCAAGAAAGCTTCAAA 77  
DB 498 TTCAAAGATTATTTTTCCTGCTGGAATACCTTTGTAGAAAACGCAAGAAAGCTTCAAA 557

QY 78 GCTTGAGAGGCTGCTGATGAAAATTCACTGCTCTCCAGACAGCTTGGGCGATCCTT 137  
DB 558 GCTTGAGAGGCTGCTGATGAAAATTCACTGCTCTCCAGACAGCTTGGGCGATCCTT 617

QY 138 TTG 140  
DB 618 TTG 620

RESULT 6  
ABS73287  
ID ABS73287 standard; DNA; 2791 BP.  
XX  
XX ABS73287;  
XX  
XX 04-DEC-2002 (first entry)  
XX  
XX DNA encoding human translocation del(12p) protein #2.  
XX  
XX  
DE Chromosome aberration; oncogenic fusion protein; cancer;  
XX proliferative disease; cellular protein isoform; heat shock protein 90;  
KM HSP-90; rheumatoid arthritis; cancer; haematopoietic disorder;  
KM T cell lymphoma; B cell lymphoma; chronic myeloid leukaemia; CML;  
KM acute myeloid leukaemia; AML; chronic myelomonocytic leukaemia; CMML;  
KM acute lymphoblastic leukaemia; ALL; APL; NHL; solid tumour;  
KM papillary thyroid carcinoma; Ewing's sarcoma; melanoma; liposarcoma;  
KM rhabdomyosarcoma; synovial sarcoma; viral infection; gene; ds.  
XX  
XX Homo sapiens.  
XX  
XX WO200269900-A2.  
XX  
XX 12-SEP-2002.  
XX  
XX 01-MAR-2002; 2002WO-US006518.  
XX  
XX 01-MAR-2001; 2001US-0272751P.  
XX  
XX (CONF-) CONFORMA THERAPEUTICS CORP.  
XX  
XX Fritz LC, Burrows FJ;  
XX  
XX WPI: 2002-698710/75.  
XX  
XX P-PSDB; ABG95083.  
XX  
XX Treating genetically-defined disease associated with chromosomal  
PT aberrations yielding oncogenic fusion proteins, e.g. cell proliferative  
PT diseases, involves administering an inhibitor of heat shock protein 90.  
XX  
XX Disclosure: Page 246-247; 389pp; English.  
XX  
XX The invention describes a method of treating genetically-defined disease  
XX associated with chromosomal aberrations yielding oncogenic fusion  
XX proteins (I), treating cancerous cells containing (I) in a heterogeneous  
XX cell population, treating proliferative diseases associated with mutant  
XX protein or cellular protein isoforms (II) dependent on heat shock protein  
XX (HSP)-90, or selectively treating cells expressing (II) involving  
XX administering HSP90-inhibitor. The method is useful for treating  
XX genetically-defined disease with chromosomal aberration yielding  
XX oncogenic fusion protein, treating cancerous cells containing fusion  
XX protein in heterogeneous cell population, treating proliferative disease  
XX (e.g. Rheumatoid arthritis or cancer) associated with mutant protein or  
XX cellular protein isoform dependent on heat shock protein (HSP)-90 (e.g.  
XX p53), or selectively treating cells expressing mutant protein or cellular  
XX protein isoform in a patient heterozygous for (II). The method is useful  
XX for treating a disease e.g. haematopoietic disorder such as T or B cell  
XX lymphoma, chronic myeloid leukaemia (CML), APL, ALL, AML, NHL and CMML,  
XX or a disease characterised by a solid tumour such as papillary thyroid  
XX carcinoma, Ewing's sarcoma, melanoma, liposarcoma, rhabdomyosarcoma and  
XX synovial sarcoma. The method is also useful for treating viral  
XX infections. This represents the DNA sequence of a chromosome aberration



SQ Sequence 2791 BP; 842 A; 548 C; 625 G; 776 T; 0 U; 0 Other;  
 Query Match 39.3%; Score 118.2; DB 6; Length 2791;  
 Best Local Similarity 97.6%; Pred. No. 2.1e-16;  
 Matches 120; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 18 TTGAGATTATTTTACTGCTGGAATATTCTTTGAGAAACCAAGAAAGAACTTCAA 77  
 497 TTCAAAGATTATTTTACTGCTGGAATATTCTTTGAGAAACCAAGAAAGAACTTCAA 556  
 DB 78 GCCTGGAGAGGCTGCATGAAATTCAGTTCGCTCTCCAGACAGCTTGCGCATCCT 137  
 DB 557 GCCTGGAGAGGCTGCATGAAATTCAGTTCGCTCTCCAGACAGCTTGCGCATCCT 616  
 QY 138 TTG 140  
 617 TTG 619  
 DB  
 RESULT 7  
 ABS73288  
 ID ABS73288 standard; DNA; 2791 BP.  
 AC ABS73288;  
 XX  
 XX 04-DEC-2002 (first entry)  
 DT  
 DE DNA encoding human translocation del(12p) protein #3.  
 XX  
 XX Chromosome aberration; oncogenic fusion protein; cancer;  
 XX proliferative disease; cellular protein isoform; heat shock protein 90;  
 XX HSP-90; rheumatoid arthritis; cancer; haematopoietic disorder;  
 XX T cell lymphoma; B cell lymphoma; chronic myeloid leukaemia; CML;  
 XX acute myeloid leukaemia; ALL; chronic myelomonocytic leukaemia; CMML;  
 XX acute lymphoblastic leukaemia; ALL; APL; NHL; solid tumour;  
 XX papillary thyroid carcinoma; Ewing's sarcoma; melanoma; liposarcoma;  
 XX rhabdomyosarcoma; synovial sarcoma; viral infection; gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200269900-A2.  
 XX  
 PD 12-SEP-2002.  
 XX  
 PF 01-MAR-2002; 2002WO-US006518.  
 PR  
 PR 01-MAR-2001; 2001US-0272751P.  
 XX  
 PA (CONF-) CONFORMA THERAPEUTICS CORP.  
 XX  
 PI Fritz LC, Burrows FU;  
 XX  
 DR P-PSDB; ABG95084.  
 XX  
 PT Treating genetically-defined disease associated with chromosomal  
 PT aberrations yielding oncogenic fusion proteins, e.g. cell proliferative  
 PT diseases, involves administering an inhibitor of heat shock protein 90.  
 XX  
 PS Disclosure; Page 248-249; 389pp; English.  
 CC The invention describes a method of treating genetically-defined disease  
 CC associated with chromosomal aberrations yielding oncogenic fusion  
 CC proteins (I), treating cancerous cells containing (I) in a heterogeneous  
 CC cell population, treating proliferative diseases associated with mutant  
 CC protein or cellular protein isoforms (II) dependent on heat shock protein  
 CC (HSP)-90, or selectively treating cells expressing (II) involving  
 CC administering HSP90-inhibitor. The method is useful for treating  
 CC genetically-defined disease with chromosomal aberration yielding  
 CC oncogenic fusion protein, treating cancerous cells containing fusion  
 CC protein in heterogeneous cell population, treating proliferative disease  
 CC (e.g. rheumatoid arthritis or cancer) associated with mutant protein or  
 CC cellular protein isoform dependent on heat shock protein (HSP)-90 (e.g.

CC p53), or selectively treating cells expressing mutant protein or cellular  
 CC protein isoform in a patient heterozygous for (II). The method is useful  
 CC for treating a disease e.g. haematopoietic disorder such as T or B cell  
 CC lymphoma, chronic myeloid leukaemia (CML), APL, ALL, ANL, NHL and CMML,  
 CC or a disease characterised by a solid tumour such as papillary thyroid  
 CC carcinoma, Ewing's sarcoma, melanoma, liposarcoma, rhabdomyosarcoma and  
 CC synovial sarcoma. The method is also useful for treating viral  
 CC infections. This represents the DNA sequence of a chromosome aberration  
 XX  
 SQ Sequence 2791 BP; 842 A; 548 C; 625 G; 776 T; 0 U; 0 Other;  
 Query Match 39.3%; Score 118.2; DB 6; Length 2791;  
 Best Local Similarity 97.6%; Pred. No. 2.1e-16;  
 Matches 120; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 18 TTGAGATTATTTTACTGCTGGAATATTCTTTGAGAAACCAAGAAAGAACTTCAA 77  
 497 TTCAAAGATTATTTTACTGCTGGAATATTCTTTGAGAAACCAAGAAAGAACTTCAA 556  
 DB 78 GCCTGGAGAGGCTGCATGAAATTCAGTTCGCTCTCCAGACAGCTTGCGCATCCT 137  
 DB 557 GCCTGGAGAGGCTGCATGAAATTCAGTTCGCTCTCCAGACAGCTTGCGCATCCT 616  
 QY 138 TTG 140  
 617 TTG 619  
 DB  
 RESULT 8  
 AAC55312  
 ID AAC55312 standard; cDNA; 2818 BP.  
 AC AAC55312;  
 XX  
 XX 05-FEB-2001 (first entry)  
 DT  
 DE Human activation-induced cytidine deaminase encoding cDNA SEQ ID NO.7.  
 XX  
 XX Activation-induced cytidine deaminase; AID; cytidine deaminase;  
 XX immune related disease; allergy; allergic disease; anti-allergic;  
 XX antianaemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;  
 XX gene therapy; B cell associated immune system disorder; food allergy;  
 XX immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
 XX IGA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;  
 XX drug allergy; allergic rhinitis; Rosen disease; Digorge disease; AIDS;  
 XX ataxia telangiectasia; common variable immunodeficiency disorder;  
 XX major histocompatibility class II deficiency disease;  
 XX auto immunodeficiency syndrome; IgG subclass selection disorder; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 80..676  
 FT /\*tag= a  
 FT /product= "activation-induced cytidine deaminase"  
 XX  
 PN WO2000058480-A1.  
 XX  
 PD 05-OCT-2000.  
 XX  
 PF 28-MAR-2000; 2000WO-JP001918.  
 XX  
 PR 29-MAR-1999; 99JP-00087192.  
 PR 24-JUN-1999; 99JP-00178899.  
 PR 27-DEC-1999; 99JP-00371182.  
 XX  
 PA (NIBS) JAPAN TOBACCO INC.  
 PA (HONJ) HONJO T.  
 XX  
 PI Honjo T, Muramatsu M;  
 XX  
 DR WPI; 2000-611715/58.  
 DR P-PSDB; AAB24198.



XX Nucleic acid encoding activation induced cytidine deaminase, useful as a  
PT target for drug development for immune-related diseases including  
PT allergies.  
PS Claim 3; Page 135-139; 174pp; Japanese.  
XX  
XX The present sequence encodes human activation-induced cytidine deaminase  
CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and has  
CC cytidine activity similar to APOBEC-1. AID has antiallergic, antianemic,  
CC antiasthmatic, ophthalmological, anti-HIV and dermatological activities,  
CC and can be used in gene therapy. AID polynucleotides are useful in  
CC methods for identifying drugs for the treatment of B cell associated  
CC immunoglobulin A (IGA) deficiency disease, IGA nephritis, gamma-  
CC globulinaemia, atopic dermatitis, allergic colitis, asthma, food allergy,  
CC drug allergy, allergic rhinitis, Rosen disease, Digorge disease, ataxia  
CC telangiectasia, common variable immunodeficiency disorder, MHC (major  
CC histocompatibility class) class II deficiency disease, AIDS (auto  
CC immunodeficiency syndrome), elevated IGE disorder, and Igg subclass  
CC selection disorder. The DNA sequences encoding AID may be used for gene  
CC therapy and the antibodies to the AID protein may be used for diagnosis  
CC and treatment of these disorders  
XX  
SQ Sequence 2818 BP; 868 A; 548 C; 626 G; 776 T; 0 U; 0 Other;  
XX  
Query Match 39.3%; Score 118.2; DB 3; Length 2818;  
Best Local Similarity 97.6%; Pred. No. 2.1e-16;  
Matches 120; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 18 TTGAGATTATTTTACTGCTGGAATACCTTTGTGAGAAACCCAGAAAGACTTTCAA 77  
DB 500 TTCAAGATTATTTTACTGCTGGAATACCTTTGTGAGAAACCCAGAAAGACTTTCAA 559  
QY 78 GCTGGAGAGGCTGCATGAAATTCAGTCTCTCCAGACAGCTTGGCGCATCCTT 137  
DB 560 GCTGGAGAGGCTGCATGAAATTCAGTCTCTCCAGACAGCTTGGCGCATCCTT 619  
QY 138 TTG 140  
DB 620 TTG 622  
XX  
AC AACS5318; standard; DNA; 116 BP.  
XX  
AC AACS5318;  
XX  
DT 05-FEB-2001 (first entry)  
XX  
DE Human activation-induced cytidine deaminase exon 4 SEQ ID NO:14.  
XX  
XX Activation-induced cytidine deaminase; AID; cytidine deaminase;  
KM immune related disease; allergy; allergic disease; antiallergic;  
KM antianemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;  
KM gene therapy; B cell associated immune system disorder; food allergy;  
KM immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
KM IGA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;  
KM drug allergy; allergic rhinitis; Rosen disease; Digorge disease; AIDS;  
KM ataxia telangiectasia; common variable immunodeficiency disorder;  
KM major histocompatibility class II deficiency disease;  
KM auto immunodeficiency syndrome; Igg subclass selection disorder; ds.  
XX  
XX Homo sapiens.  
XX  
PN WO200058480-A1.  
XX  
XX 05-OCT-2000.  
XX  
XX 28-MAR-2000; 2000MO-JP001918.  
XX  
XX 29-MAR-1999; 99JP-00087192.

PR 24-JUN-1999; 99JP-00178999.  
PR 27-DEC-1999; 99JP-00371382.  
XX  
XX (NIBS) JAPAN TOBACCO INC.  
PA (HONU/) HONJO T.  
XX  
XX Honjo T, Muramatsu M;  
PI WPI; 2000-611715/58.  
XX  
DR Nucleic acid encoding activation induced cytidine deaminase, useful as a  
PT target for drug development for immune-related diseases including  
PT allergies.  
XX  
XX Claim 18; Page 151; 174pp; Japanese.  
XX  
PS The present invention describes an activation-induced cytidine deaminase  
CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and has  
CC cytidine activity similar to APOBEC-1. AID has antiallergic, antianemic,  
CC antiasthmatic, ophthalmological, anti-HIV and dermatological activities,  
CC and can be used in gene therapy. AID polynucleotides are useful in  
CC methods for identifying drugs for the treatment of B cell associated  
CC immune system disorders, immunodeficiency diseases and allergies, such as  
CC immunoglobulin A (IGA) deficiency disease, IGA nephritis, gamma-  
CC globulinaemia, atopic dermatitis, allergic colitis, asthma, food allergy,  
CC drug allergy, allergic rhinitis, Rosen disease, Digorge disease, ataxia  
CC telangiectasia, common variable immunodeficiency disorder, MHC (major  
CC histocompatibility class) class II deficiency disease, AIDS (auto  
CC immunodeficiency syndrome), elevated IGE disorder, and Igg subclass  
CC selection disorder. The DNA sequences encoding AID may be used for gene  
CC therapy and the antibodies to the AID protein may be used for diagnosis  
CC and treatment of these disorders. The present sequence represents the  
CC exon 4 genomic DNA sequence of human AID  
XX  
SQ Sequence 116 BP; 32 A; 25 C; 24 G; 35 T; 0 U; 0 Other;  
XX  
Query Match 38.5%; Score 116; DB 3; Length 116;  
Best Local Similarity 100.0%; Pred. No. 4.4e-16;  
Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 25 ATTATTTTACTGCTGGAATACCTTTGTGAGAAACCCAGAAAGACTTTCAAAGCCTGG 84  
DB 1 ATTATTTTACTGCTGGAATACCTTTGTGAGAAACCCAGAAAGACTTTCAAAGCCTGG 60  
QY 85 AAGGCTGCATGAAATTCAGTCTCTCCAGACAGCTTGGCGCATCCTTTG 140  
DB 61 AAGGCTGCATGAAATTCAGTCTCTCCAGACAGCTTGGCGCATCCTTTG 116  
XX  
AC AACS5307;  
XX  
DT 05-FEB-2001 (first entry)  
XX  
DE Mouse activation-induced cytidine deaminase encoding cDNA SEQ ID NO:1.  
XX  
XX Activation-induced cytidine deaminase; AID; cytidine deaminase;  
KM immune related disease; allergy; allergic disease; antiallergic;  
KM antianemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;  
KM gene therapy; B cell associated immune system disorder; food allergy;  
KM immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
KM IGA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;  
KM drug allergy; allergic rhinitis; Rosen disease; Digorge disease; AIDS;  
KM ataxia telangiectasia; common variable immunodeficiency disorder;  
KM major histocompatibility class II deficiency disease;  
KM auto immunodeficiency syndrome; Igg subclass selection disorder; ss.  
XX  
XX Mus musculus.  
XX  
XX  
XX Key Location/Qualifiers



```
FT CDS 93..689
FT /*tag= a
FT /product= "activation-induced cytidine deaminase"
XX XX
XX PN MO200058480-A1.
XX
XX PD 05-OCT-2000.
XX
XX PF 28-MAR-2000; 2000WO-JP001918.
XX
XX PR 29-MAR-1999; 99JP-00087192.
XX PR 24-JUN-1999; 99JP-00178999.
XX PR 27-DEC-1999; 99JP-00371382.
XX
XX PA (NIBS ) JAPAN TOBACCO INC.
XX PA (HONJ/) HONJO T.
XX
XX PI Honjo T, Muramatsu M;
XX
XX PS WPI; 2000-611715/58.
XX DR F-PSDB; AAB24197.
XX
XX PT Nucleic acid encoding activation induced cytidine deaminase, useful as a
XX PT target for drug development for immune-related diseases including
XX PT allergies.
XX
XX PS Claim 3; Page 126-130; 174pp; Japanese.
XX
XX CC The present sequence encodes mouse activation-induced cytidine deaminase
XX CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and has
XX CC cytidine activity similar to APOBEC-1. AID has antiallergic, antitumoric,
XX CC antiasthmatic, ophthalmological, anti-HIV and dermatological activities,
XX CC and can be used in gene therapy. AID polynucleotides are useful in
XX CC methods for identifying drugs for the treatment of B cell associated
XX CC immune system disorders, immunodeficiency diseases and allergies, such as
XX CC immunoglobulin A (IGA) deficiency disease, IGA nephritis, gamma-
XX CC globulinemia, atopic dermatitis, allergic colitis, asthma, food allergy,
XX CC drug allergy, allergic rhinitis, Rosen disease, DiGeorge disease, ataxia
XX CC telangiectasia, common variable immunodeficiency disorder, MHC (major
XX CC histocompatibility class II deficiency disease, AIDS (auto
XX CC immunodeficiency syndrome), elevated IgG disorder, and IgG subclass
XX CC selection disorder. The DNA sequences encoding AID may be used for gene
XX CC therapy and the antibodies to the AID protein may be used for diagnosis
XX CC and treatment of these disorders
XX
XX SQ Sequence 2440 BP; 706 A; 546 C; 551 G; 636 T; 0 U; 1 Other;
XX
XX Query Match 33.4%; Score 100.6; DB 3; Length 2440;
XX Best Local Similarity 88.6%; Pred. No. 1.3e-12;
XX Matches 109; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
XX
XX QY 18 TTTGAAGATATTTTACTGCTGGAATCTTTGTGAAACACAGAAAGACTTTCAA 77
XX DB 513 TTCAAAGCTATTTTACTGCTGGAATCATTTGTGAAACATGTGAAAGACTTTCAA 572
XX
XX QY 78 GCGTGGAGAGGCGCATGAAATTCAGTTCCTCTCCAGACAGCTTCGGCGATCCTT 137
XX DB 573 GCGTGGAGAGGCGCATGAAATTCCTCGGCTAACCAACACTTCGGCGATCCTT 632
XX
XX QY 138 TTG 140
XX |||
XX DB 633 TTG 635
XX
XX RESULT 11
XX AA187261/c
XX ID AA187261 standard; cDNA; 326 BP.
XX
XX AC AA187261;
XX
XX XX 06-NOV-2001 (first entry)
XX
XX DE Human polynucleotide SEQ ID NO 7321.
```

```
XX XX
XX KM Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX KM tissue growth factor; immunomodulatory; cancer; leukaemia;
XX KM nervous system disorders; arthritis; inflammation; ss.
XX
XX OS Homo sapiens.
XX
XX PN MO200164835-A2.
XX
XX PD 07-SEP-2001.
XX
XX PF 26-FEB-2001; 2001WO-US004927.
XX
XX PR 28-FEB-2000; 2000US-00515126.
XX PR 18-MAY-2000; 2000US-00577409.
XX
XX PA (HYSE-) HYSEQ INC.
XX
XX PI Tang YT, Liu C, Drmanac RT;
XX
XX XX WPI; 2001-514838/56.
XX DR P-PSDB; AA007330.
XX
XX PT Isolated nucleic acids and polypeptides, useful for preventing diagnosing
XX PT and treating e.g. leukemia, inflammation and immune disorders.
XX
XX PS Claim 1; SEQ ID NO 7321; 1399pp + Sequence Listing; English.
XX
XX CC The invention relates to human polynucleotides (AA179941-AA193841) and
XX CC the encoded proteins (AA000010-AA013910) that exhibit activity relating to
XX CC cytokine, cell proliferation or cell differentiation or which may induce
XX CC production of other cytokines in other cell populations. The
XX CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX CC peptide therapy. The polypeptides have various cytokine-like activities,
XX CC e.g. stem cell growth factor activity, haematopoiesis regulating
XX CC activity, tissue growth factor activity, immunomodulatory activity and
XX CC activity/inhibit activity and may be useful in the diagnosis and/or
XX CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX CC inflammation. Note: The sequence data for this patent did not form part
XX CC of the printed specification, but was obtained in electronic format
XX CC directly from WIPO at ftp.wipo.int/pub/published_pat_sequences
XX
XX SQ Sequence 326 BP; 204 A; 23 C; 42 G; 57 T; 0 U; 0 Other;
XX
XX Query Match 16.7%; Score 50.4; DB 4; Length 326;
XX Best Local Similarity 49.6%; Pred. No. 0.082;
XX Matches 129; Conservative 0; Mismatches 131; Indels 0; Gaps 0;
XX
XX QY 11 TTTTGTGGAAGATTTTACTGCTGGAATCTTTGTGAAACACAGAAAGACTTTCAA 70
XX DB 278 TTTTGTGGAAGATTTTACTGCTGGAATCTTTGTGAAACACAGAAAGACTTTCAA 219
XX
XX QY 71 TTCAAAGCTATTTTACTGCTGGAATCATTTGTGAAACATGTGAAAGACTTTCAA 572
XX DB 218 TTGCGGGAATTTTGTGCGGTTTAAATAATTTTATTTTTCCTCCCGGGGG 159
XX
XX QY 131 CATCCTTTGGTAAGGCGCTCTCGCTTTAAATTTCTTCTTCACAGTCTT 190
XX DB 158 TTTTGTGGAAGATTTTACTGCTGGAATCTTTGTGAAACACAGAAAGACTTTCAA 70
XX
XX QY 191 TTTGAGTTTGGTAATTTCTATATTTCTATGTCATCATCTCAGTTCATCT 250
XX DB 98 TTTTGTGGAAGATTTTACTGCTGGAATCATTTGTGAAACATGTGAAAGACTTTCAA 572
XX
XX QY 251 GATGAAACTTTATTTCTCC 270
XX DB 38 AATTAATAGTTTATTTATTC 19
XX
XX RESULT 12
XX ABV55592/c
XX ID ABV55592 standard; cDNA; 490 BP.
XX
```



[illegible]

XX	ADBS54224 standard; DNA; 4316 BP.
XX	ID ADBS4224
AC	ADB54224;
XX	
DT	04-DEC-2003 (first entry)
XX	
DE	Pretreated genomic DNA region 148.
XX	
KM	colon cell proliferative disorder; non methylated CpG dinucleotide;
XX	cytostatic; cancer; adenoma; carcinoma; cytosine methylation state; ds.
OS	Unidentified.
XX	
FN	NC02003072821-A2.
XX	
PD	04-SEP-2003.
XX	
PF	27-FEB-2003; 2003WO-EP002035.
XX	
PR	27-FEB-2002; 2002EP-00004551.
PA	(EPIG-) EPIGENOMICS AG.
XX	
FI	Adorjan P, Burger M, Maier S, Nimmrich I, Becker E, Lesche R,
PI	Rujan T, Schmitt A;
XX	
DR	WPI; 2003-731620/69.
XX	
PT	Detecting and differentiating between colon cell proliferative disorders
XX	associated with a gene or its regulatory regions comprises contacting a
PT	target nucleic acid in a biological sample obtained from the subject with
PT	a reagent.
PS	Claim 32; SEQ ID NO 280; 74pp; English.
XX	
CC	The invention relates to a novel method for detecting and differentiating
CC	between colon cell proliferative disorders associated with at least one
CC	gene or its regulatory regions. The method comprises contacting a target
CC	nucleic acid in a biological sample obtained from the subject with at
CC	least one reagent or a series of reagents, where the reagent or series of
CC	reagents, distinguishes between methylated and non methylated CpG
CC	dinucleotides within the target nucleic acid. The molecules of the
CC	invention demonstrate cytostatic activity whilst the method may useful
CC	for detecting and differentiating between colon cell proliferative
CC	disorders, including cancers such as colon adenoma and colon carcinoma.
CC	The pNA (peptide nucleic acid)-oligomers are useful as probes for
CC	determining cytosine methylation state or single nucleotide
CC	polymorphisms. The current sequence is that of the pretreated genomic DNA
CC	region of the invention. This sequence is not shown within the
CC	specification but is taken from Wipweb.
XX	
SQ	Sequence 4316 BP; 934 A; 0 C; 1090 G; 2292 T; 0 U; 0 Other;
XX	
Query Match	16.1%; Score 48.4; DB 9; Length 4316;
Best Local Similarity	47.7%; Pred. No. 0.3; Mismatches 156; Indels 0; Gaps 0;
Matches 142; Conservative	0; Mismatches 156; Indels 0; Gaps 0;
OY	1 TTTTCCTTTTTTTTTTTTTTGAAGATTATTTTACTGTGGAACTTTGTAGAAAAC 60
Db	3890 TTTTAGTTTTTGATGTAATTTTTTGTTTTTTATTTAATTTGATTTTTTGATTTT 3944
OY	61 ACAGAAAGACTTCAAACCCTGGAGAAGCGCTGCATGAATAATTCAGTTCCTCAAGC 120
Db	3950 TTGGAGAGAAATTTATTTTAAGAGGTTGGTGTTAATGATATGATTTGTTGAGAAAGA 4008
OY	121 AGCTTGCGGAGATCCTTTGGTAAAGGCGCTCTCGCTTTTAAATTTCTTTCTTC 180
Db	4010 AGGTTTAAGATAGTT 4066
OY	181 TACAGCTTTTTTGGAGTTTGGTATATTTCTATTTCTATTTGTTCAATCACTCTCA 240







SQ Sequence 506 BP; 195 A; 101 C; 95 G; 107 T; 0 U; 8 Other;

Query Match 15.9%; Score 47.8; DB 8; Length 506;

Best Local Similarity 58.6%; Pred. No. 0.32; Mismatches 58; Indels 0; Gaps 0;

Matches 82; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 127 GGGGATCCTTTTGGTAAGGGGCTTCGCTTTTAAATTTCTTTCTCTACAGT 186

Db 324 GGGCCTTTTATTTTAAAGGGGGGTTTTTTTTTTTTTTTTTTTTTTTTTTTT 265

QY 187 CTTTTTGAAGTTTGGTAATTTCTTAATTTCTTAATTTGTAATCACTCTGAGTTTC 246

Db 264 TT 205

QY 247 ATCTGATGAAGAACTTATTT 266

Db 204 GTTCCAATTAAGAACTTATTT 185

Search completed: March 12, 2004, 18:08:04  
Job time : 123.53 secs















of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences. Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as low coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

## QUALSTAR-REPORT:

## FEATURES

## source

Location/Qualifiers  
1. 71132  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="12"  
/clone="RP11-438L7"  
/complement(1..1448)  
/note="overlaps bases 1..1448 of clone AC092490"  
/function="clone overlap"  
187..408  
/standard\_name="57233"  
439..560  
/standard\_name="92005"  
774..881  
/rpt\_family="MIR"  
905..1190  
/rpt\_family="AluSq"  
1191..1213  
/rpt\_family="AT\_rich"  
1591..1807  
/standard\_name="6612"  
1744..1819  
/standard\_name="8198"  
1966..2264  
/rpt\_family="AluSq"  
3286..3328  
/rpt\_family="(TTTC)n"  
4097..4249  
/rpt\_family="AluSq"  
4360..4654  
/rpt\_family="AluY"  
complement(5140..5262)  
/rpt\_family="FLAM\_C"  
5275..5304  
/rpt\_family="AT\_rich"  
5305..5563  
/rpt\_family="AluO"  
5564..5599  
/rpt\_family="AT\_rich"  
5644..5733  
/rpt\_family="MSTD"  
5734..6012  
/rpt\_family="AluSq"  
6020..6336  
/rpt\_family="AluSx"  
6337..6629

/rpt\_family="MSTD"  
6330..6650  
/rpt\_family="(CAA)n"  
complement(6711..6817)  
/rpt\_family="MIR"  
repeats(7033..7327)  
/rpt\_family="AluSx"  
7607..7745  
/rpt\_family="FLAM\_A"  
7980..8047  
/rpt\_family="MIR"  
repeats(8050..8549)  
/rpt\_family="LTR47A"  
8567..8698  
/rpt\_family="MIR"  
complement(8815..9101)  
/rpt\_family="AluSq"  
complement(11227..12175)  
/rpt\_family="LTR5"  
12625..12900  
/standard\_name="87432"  
complement(13547..13664)  
/rpt\_family="LIMB5"  
complement(13665..13979)  
/rpt\_family="AluY"  
complement(13980..14193)  
/rpt\_family="LIMB5"  
14622..14924  
/rpt\_family="AluY"  
complement(15813..17136)  
/rpt\_family="LIPAB"  
complement(17272..17395)  
/rpt\_family="FLAM\_C"  
17426..17575  
/rpt\_family="MIR"

Query Match 98.9%; Score 297.8; DB 9; Length 71132;

Best Local Similarity 99.3%; Pred. No. 1,1e-49; Matches 299; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTTTCCTTTTGTGAGATTATTTTACGCTGGAATCTTTGTGAAAAC 60  
DB 42890 TTTTCCTTTTGTGAGATTATTTTACGCTGGAATCTTTGTGAAAAC 42949  
QY 61 ACGAAAGACTTTCAAGAGCTGGAGAGGCTGATGAAATAGTTCCTCCAGAC 120  
DB 42950 ACGAAAGACTTTCAAGAGCTGGAGAGGCTGATGAAATAGTTCCTCCAGAC 43009  
QY 121 AGCTTGCGGCGATCCTTTGTAAGGGGCTTCGCTTTTAATTTCTTCTTCTC 180  
DB 43010 AGCTTGCGGCGATCCTTTGTAAGGGGCTTCGCTTTTAATTTCTTCTTCTC 43069  
QY 181 TACAGCTTTTGGAGTTTCGTATATTTCTTATTTTCTTATTTGTCATGACTCTCA 240  
DB 43070 TACAGCTTTTGGAGTTTCGTATATTTCTTATTTTCTTATTTGTCATGACTCTCA 43129  
QY 241 GTTTTCATCTGATGAAGACTTATTTCTCTCCATCAGACTTTTCTTCTGCTGTCA 300  
DB 43130 GTTTTCATCTGATGAAGACTTATTTCTCTCCATCAGACTTTTCTTCTGCTGTCA 43189  
QY 301 C 301  
DB 43190 C 43190

RESULT 5  
HSA577811 HSA577811 335 bp mRNA linear PRI 22-JUL-2003  
LOCUS Homo sapiens partial mRNA for activation-induced cytidine deaminase  
DEFINITION (AID gene).  
ACCESSION A577811  
VERSION A577811.1 GI:33145978  
KEYWORDS activation-induced cytidine deaminase; AID gene.



SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS 1  
TAKHAR, P., COKE, H.A., FEAR, D., SMURTHWAITE, L., DURHAM, S.R. and  
GOULD, H.J.  
TITLE Allergen and activation-induced deaminase drive heavy-chain class switching to IgE in the nasal mucosa of hay fever patients  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 335)  
AUTHORS Takhar, P  
TITLE Direct Submission  
JOURNAL Submitted (21-JUL-2003) Takhar P., Biomedical Sciences Randall Centre, King's College London, New Hunt House, Guy's Campus, St Thomas St, London, SE1 1UL, UNITED KINGDOM  
FEATURES  
source  
1.335  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/isolation\_source="nasal biopsy from inferior turbinate of an allergic rhinitis patient"  
/db\_xref="taxon:9606"  
/clone="NR02"  
/sex="female"  
/cell\_type="B lymphocyte"  
/tissue\_type="nasal mucosa"  
/dev\_stage="adult"  
/country="United Kingdom"  
1.335  
/gene="AID"  
/cDS  
1.335  
/function="editing (involved in somatic hypermutation and class switch recombination)"  
/evidence="experimental"  
/product="activation-induced cytidine deaminase"  
/protein\_id="CAI2261.1"  
/db\_xref="GI:33145979"  
/translation="DPGRCYRVTFWTSNPPCDCAHVADFLRGNNPISLRIFETARLY FCDRAEPGLRLHRAVGQIAIMTFKDYFCWTFVENHRTFKAMGGLHNSVRL SROLRRILL"  
ORIGIN  
Query Match 39.8%; Score 119.8; DB 9; Length 335;  
Best Local Similarity 98.4%; Pred. No. 5.6e-14;  
Matches 121; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 18 TTTGAAGATTATTTTACGCTGGAATACCTTTGTAGAAAACGAGAAAGAACTTTCAA 77  
DB 213 TTTCAAGATTATTTTACGCTGGAATACCTTTGTAGAAAACGAGAAAGAACTTTCAA 272  
QY 78 GCCTGGAGAGGCTGCATGAATTCAGTTCCTCTCCAGACAGCTTCGGCGCATCCTT 137  
DB 273 GCCTGGAGAGGCTGCATGAATTCAGTTCCTCTCTCCAGACAGCTTCGGCGCATCCTT 332  
QY 138 TTG 140  
DB 333 TTG 335  
RESULT 6  
AF529837 596 bp mRNA linear SYN 17-SEP-2002  
LOCUS Mus musculus clone 10 transgenic Homo sapiens AID (AID) mRNA,  
DEFINITION complete cds.  
ACCESSION AF529837  
VERSION AF529837.1 GI:22297261  
KEYWORDS Mus musculus (house mouse)  
SOURCE Mus musculus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
AUTHORS 1 (bases 1 to 596)  
MARTIN, A. and SCHARFF, M.D.  
TITLE Somatic hypermutation of the AID transgene in B and non-B cells  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12304-12308 (2002)  
REFERENCE 2 (bases 1 to 596)  
AUTHORS Martin, A. and Scharff, M.D.  
TITLE Direct Submission  
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA  
FEATURES  
source  
1.596  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="110"  
/cell\_type="hybridoma PL-5"  
/transgenic  
1.596  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
1.596  
/gene="AID"  
1.596  
/gene="AID"  
/codon\_start=1  
/product="AID"  
/protein\_id="PAM95424.1"  
/db\_xref="GI:22297262"  
/translation="MDSLMMRRKFLYQFNWVWAKGRRETYLCVYVRRDSATSPSL DFGYLRNKGCHVELFLRYISDWDLPBGCVRYVTFWTSNPPCDCAHVADFLRGNP NLSLRIFETARLYFCDRAEPGLRLHRAVGQIAIMTFKDYFCWTFVENHRTFK AMGLHNSVRLSROLRRILLPLRYVDLDAFRTTGL"  
ORIGIN  
Query Match 39.6%; Score 119.2; DB 12; Length 596;  
Best Local Similarity 97.6%; Pred. No. 6.7e-14;  
Matches 121; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 17 TTTGAAGATTATTTTACGCTGGAATACCTTTGTAGAAAACGAGAAAGAACTTTCAA 76  
DB 420 TTTCAAGATTATTTTACGCTGGAATACCTTTGTAGAAAACGAGAAAGAACTTTCAA 479  
QY 77 AGCTGGAGAGGCTGCATGAATTCAGTTCCTCTCCAGACAGCTTCGGCGCATCCT 136  
DB 480 AGCTGGAGAGGCTGCATGAATTCAGTTCCTCTCTCCAGACAGCTTCGGCGCATCCT 539  
QY 137 TTG 140  
DB 540 TTG 543  
RESULT 7  
AC094826/c 241757 bp DNA linear HTG 09-MAY-2003  
LOCUS Rattus norvegicus clone CH230-2013, WORKING DRAFT SEQUENCE.  
ACCESSION AC094826  
VERSION AC094826.4 GI:30466489  
KEYWORDS HTG; HTGS PHASE2; HTGS DRAFT; HTGS FULLTOP.  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
REFERENCE 1 (bases 1 to 241757)  
MURPHY, D., MARIE, Metzker, M., LEE, Abramson, S., ADAMS, C., ALDER, J., ALLEN, C., ALLEN, H., ALSTROCK, S., AMIN, A., ANGUSANO, D., ANGLEBECH, V., AYOGLI, A., AYODEJI, M., BACA, E., BADEN, H., BALDWIN, D., BANDARANAIAKE, D., BARBER, M., BARNSTEAD, M., BENAHMED, F., BIEWALD, K., BLAIR, J., BLANKENBURY, K., BLYTH, P., BROWN, M., BRYANT, N., BUHAY, C., BURCH, P., BURRELL, K., CALDERON, E., CARDENAS, V., CARTER, K., CAVAZOS, I., CEASAR, H., CENTER, A.,



[illegible]







KEYWORDS  
SOURCE  
ORGANISM

HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FUILLTOP.  
Rattus norvegicus (Norway rat)  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

REFERENCE  
AUTHORS

1 (bases 1 to 345098)  
Muzny,D,Marie, Metzker,M,lee, Abramzon,S, Adams,C, Alder,J, Allen,C, Allen,R, Alsbrooks,S, Amin,A, Anguiano,D, Anyalebech,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H, Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benahmed,F, Biswal,K, Blair,J, Blankenburg,K, Blyth,P, Brown,M, Bryant,N, Bunay,C, Burch,P, Burrell,K, Calderon,E, Cardenas,V, Carter,K, Cavazos,I, Caesar,H, Center,A, Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J, Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L, Davila,M,L, Davis,C, Davy-Carroll,L, De Anda,C, Dederich,D, Delgado,O, Denson,S, Deramo,C, Ding,Y, Dim,H, Divya,K, Draper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Evans,K, Bgan,A, Escotto,M, Eugene,C, Evans,C,A, Falls,T, Fan,G, Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P, Frazer,C,M, Gabisi,A, Ganta,R, Garcia,A, Garner,T, Garza,M, Gebregeorgis,E, Geer,K, Gill,R, Grady,M, Guerra,M, Guevara,M, Gunaratne,P, Haaland,M, Hamill,C, Hamilton,C, Hamilton,K, Harvey,Y, Havlak,P, Hawes,A, Henderson,N, Hernandez,J, Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hogues,M, Hollins,B, Howells,S, Hulik,S, Hume,J, Idlebird,D, Jackson,A, Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jolivet,A, Karpathy,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kovar,C, Kowis,C, Kraft,C,L, Lebow,H, Levan,J, Lewis,L, Li,Z, Liu,J, Liu,J, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,J, Lorensheva,L, Louisedge,H, Lozano,R,J, Lu,X, Ma,J, Maheshwari,M, Mahindratne,M, Mahmoud,M, Malloy,K, Mangum,A, Mangun,B, Mapua,P, Martin,K, Martin,R, Martinez,E, Mawhinney,S, McLeod,M,P, McNeill,T,Z, Meenen,E, Mloaavljovic,A, Miner,G, Minja,E, Montemayor,J, Moore,S, Morgan,M, Morris,K, Morris,S, Mundassa,M, Murphy,M, Natr,L, Nankervis,C, Neal,D, Newton,N, Nguyen,N, Norris,S, Nnaolelehen,O, Okunomu,G, Olarinmugbon,A, Pal,S, Parks,K, Pasternak,S, Paul,H, Perez,A, Perez,L, Plankoch,C, Plopper,F, Polidexter,A, Popovic,D, Primus,E, Pu,L, L, Pazo,M, Quiroz,J, Rachlin,E, Reeves,K, Regier,M,A, Reigh,R, Rellly,B, Reilly,M, Ren,Y, Reuter,M, Richards,S, Riggs,F, Rivers,C, Rodkey,T, Rojas,A, Rose,M, Rose,R, Ruiz,S,J, Sanders,M, Saverly,G, Scherer,S, Scott,G, Shatsman,S, Shen,H, Shetty,J, Shvartsbeyn,A, Sisson,I, Sitter,C,D, Sma,S,D, Sneed,A, Sodergren,E, Song,X-Z, Sorelle,R, Sosa,J, Steidle,M, Strong,R, Sutton,A, Svatek,A, Tabor,P, Taylor,C, Taylor,T, Thomas,N, Thomas,S, Tingey,A, Trejos,Z, Uemami,K, Valas,R, Vera,V, Villaseana,D, Waldron,L, Walker,B, Wang,J, Wang,Q, Wang,S, Warren,J, Warren,R, Wei,X, White,F, Williams,G, Willson,R, Wleczky,R, Woodden,H, Wortley,K, Wright,D, Wright,R, Wu,J, Yakub,S, Yen,J, Yoon,L, Yoon,V, Yu,F, Zhang,J, Zhou,J, Zhou,X, Zhao,S, Dunn,D, von Niederhausen,A, Weiss,R, Smith,D,R, Holt,R,A, Smith,H,O, Weinstock,G, and Gibbs,R.A.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Unpublished  
2 (bases 1 to 345098)  
Worley,K.C.

Direct Submission  
Submitted (09-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 345098)  
Rat Genome Sequencing Consortium.

Direct Submission  
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

## COMMENT

On Nov 15, 2002 this sequence version replaced gi:23265477.  
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas

(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GWGN  
Center clone name: CH230-3215  
----- Summary Statistics  
Assembly program: PHRAP version 0.990329  
Consensus quality: 19191 bases at least Q40  
Consensus quality: 194508 bases at least Q30  
Consensus quality: 195878 bases at least Q20  
Estimated insert size: 227256; sum-of-contigs estimation  
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

----- NOTE: Estimated insert size may differ from sequence length  
(see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 4 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 330403: contig of 330403 bp in length  
\* 330404 330503: gap of unknown length  
\* 330504 331603: contig of 1100 bp in length  
\* 331604 331703: gap of unknown length  
\* 331704 337738: contig of 6035 bp in length  
\* 337739 337838: gap of unknown length  
\* 337839 345098: contig of 7260 bp in length.  
\* Location/Qualifiers

source  
1. 345098  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"

misc\_feature  
/db\_xref="taxon:10116"  
/clone="CH230-3215"  
/complement(1177532..178431)  
/note="clone boundary  
clone\_end:3p6"

misc\_feature  
end\_sequence:RH328475"  
179290..181054  
/note="wgs end extension  
clone\_end:3p6"

## ORIGIN

Query Match 39.4%; Score 118.6; DB 2; Length 345098;  
Best Local Similarity 78.5%; Pred. No. 2.8e-14;  
Matches 142; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 1 TTTTCCTTTTCTTTTCTTTTGAAGATTTTTCCTGGAATATCTTTGAGAAAC 60  
DB 141574 TTATGCATTATTTCTCTCTTAAGACATTTTATCTGGAATACATTTGAGAAATC 141515  
QY 61 ACGAAGAACTTTAAAGCCTGGAAGGCTGATGAAATTCAGTTCGTCAGAC 120  
DB 141514 ATGAAGAACTTTAAAGCCTGGAAGGCTGATGAAATTCGTCAGTTCAGAC 141455  
QY 121 AGCTTCGCGCATCCTTTTGTGAAGGGGCTTCCTGCTTTTAAATTTCTTTTCTC 180











/product="truncated AID"  
/protein\_id="AA095405.1"  
/db\_xref="GI:22297224"  
/translation="MDSLMMRRKFLYQFKNVMAKRRRTYLCYVKKRDSATSFSL  
DFGLRNKNGCHVELLFRLYISDWLDPGCYRVWTFSPCYDCAHVAADFLRNP  
NLALRIFPARLYFCEDEHKAPEGRRLRRAHRAVQVIAIMTFEDYFCNMTFVENHERTEK  
AMEGLHNSVRLSPQLRRILLPLYEVDL"

## ORIGIN

Query Match 39.3%; Score 118.2; DB 9; Length 596;  
Best Local Similarity 97.6%; Pred. No. 1.1e-13;  
Matches 120; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 18 TTGAAGATTATTTTCTGCTGGAATCTTTGTAGAAAACGACGAAGACTTCAAA 77  
|||  
Db 421 TTCAAAGATTATTTTCTGCTGGAATCTTTGTAGAAAACGACGAAGACTTCAAA 480  
|||  
QY 78 GCTGGGAAGGGCTGCATGAAATTCAGTCTCTCCAGACAGCTTCGGCGCATCCTT 137  
|||  
Db 481 GCTGGGAAGGGCTGCATGAAATTCAGTCTCTCTCCAGACAGCTTCGGCGCATCCTT 540  
|||  
QY 138 TTG 140  
|||  
Db 541 TTG 543

## RESULT 15

AF529819

LOCUS AF529819 596 bp mRNA linear PRI 17-SEP-2002  
DEFINITION Homo sapiens clone Ramos 5 AID (AID) mRNA, partial cds.

ACCESSION

AF529819

AF529819.1 GI:22297225

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1..596

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="Ramos 5"

1..&gt;596

/gene="AID"

1..&gt;596

/gene="AID"

/note="Integrated into Burkitt's lymphoma cell line Ramos"

/codon\_start=1

/product="AID"

/protein\_id="AA095406.1"

/db\_xref="GI:22297226"

/translation="MDSLMMRRKFLYQFKNVMAKRRRTYLCYVKKRDSATSFSL  
DFGLRNKNGCHVELLFRLYISDWLDPGCYRVWTFSPCYDCAHVAADFLRNP  
NLALRIFPARLYFCEDEHKAPEGRRLRRAHRAVQVIAIMTFEDYFCNMTFVENHERTEK  
AMEGLHNSVRLSPQLRRILLPLYEVDLDAFRTGL"

ORIGIN

Query Match 39.3%; Score 118.2; DB 9; Length 596;

Best Local Similarity 97.6%; Pred. No. 1.1e-13;

Matches 120; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 18 TTGAAGATTATTTTCTGCTGGAATCTTTGTAGAAAACGACGAAGACTTCAAA 77

|||

|||

|||

|||

|||

|||

|||

|||

|||

|||

|||

Db 421 TTCAAAGATTATTTTCTGCTGGAATCTTTGTAGAAAACGACGAAGACTTCAAA 480  
|||  
QY 78 GCTGGGAAGGGCTGCATGAAATTCAGTCTCTCCAGACAGCTTCGGCGCATCCTT 137  
|||  
Db 481 GCTGGGAAGGGCTGCATGAAATTCAGTCTCTCTCCAGACAGCTTCGGCGCATCCTT 540  
|||  
QY 138 TTG 140  
|||  
Db 541 TTG 543

Search completed: March 12, 2004, 21:22:25  
Job time : 949.629 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 12, 2004, 13:19:34 ; Search time 958.46 Seconds  
(without alignments)  
9378.080 Million cell updates/sec

Title: US-09-966-880A-10\_COPY\_6264\_6564

Perfect score: 301

Sequence: 1 gcatcaaggagcactagta.....cttcacaaggcggtcaag 301

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27133289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
EST:  
1: em\_estba:\*  
2: em\_esthm:\*  
3: em\_estim:\*  
4: em\_estmu:\*  
5: em\_estrov:\*  
6: em\_estrpl:\*  
7: em\_estro:\*  
8: em\_estro:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hrc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_ges\_hum:\*  
18: em\_ges\_inv:\*  
19: em\_ges\_pln:\*  
20: em\_ges\_vrt:\*  
21: em\_ges\_fun:\*  
22: em\_ges\_mam:\*  
23: em\_ges\_mus:\*  
24: em\_ges\_pro:\*  
25: em\_ges\_pig:\*  
26: em\_ges\_vrl:\*  
27: em\_ges\_vrl:\*  
28: gb\_ges1:\*  
29: gb\_ges2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	79.6	26.4	187	9	AA249004 12585.seq
2	75.2	25.0	476	28	AA671876 HS_5462_A
3	74.2	24.7	674	29	AG051626 Pan trog1
4	74	24.6	894	13	BQ103839 ESTCERDR2

Result No.	Score	Query Match	Length	ID	Description
5	71.2	23.7	630	14	H59882 Yr04b09.r1
6	70.4	23.4	617	9	AL038727
7	70.4	23.4	651	12	BM983994
8	70.4	23.4	783	12	BG741476
9	68.8	22.9	303	10	AM948684 QV0-FT000
10	68.8	22.9	548	13	BM362298
11	68.8	22.9	737	13	BM362298
12	68.8	22.9	758	13	BM362298
13	68.8	22.9	913	13	BM362298
14	68.8	22.9	930	13	BM362298
15	67.8	22.5	699	28	AC420050
16	67.8	22.5	446	10	BF760170
17	66.8	22.2	360	28	AC012711
18	66.8	22.2	461	28	AC078850
19	66.8	22.2	1201	9	AL573148
20	66.4	22.1	443	9	AA292629
21	66.4	22.1	682	12	BC538032
22	65.2	21.7	454	28	AC322538
23	65.2	21.7	813	14	CB312335
24	65.2	21.7	721	10	AM954122
25	65	21.6	741	12	BM674784
26	64.6	21.5	311	28	B43977
27	64.6	21.5	440	28	AC734239
28	64.2	21.3	434	28	AC737126
29	64	21.3	893	13	BU542546
30	63.8	21.2	339	14	T87005
31	63.8	21.2	629	13	BM111369
32	63.8	21.2	690	28	AC028551
33	63.8	21.2	366	12	BT789069
34	63	20.9	589	12	BT789069
35	62	20.6	305	9	AA385457
36	62	20.6	401	9	AA125271
37	62	20.6	408	9	AA281997
38	62	20.6	452	9	AI379101
39	62	20.6	509	9	AI827559
40	62	20.6	913	13	BQ883450
41	61.8	20.5	674	29	AG182249
42	61.4	20.4	346	28	AC012867
43	61.2	20.3	421	28	AC016082
44	61.2	20.3	608	29	AG161212
45	61.2	20.3	861	14	CD516515

## ALIGNMENTS

RESULT 1  
AA249004/c 187 bp mRNA EST 11-MAR-1997  
LOCUS 12585.seq.F Human fetal heart, lambda ZAP Express Homo sapiens cDNA  
DEFINITION 5', mRNA sequence.  
ACCESSION AA249004 GI:1880979  
VERSION AA249004.1  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 187)  
AUTHORS Liew C.C.  
TITLE CDNA from human fetal heart (1997)  
JOURNAL Unpublished (1997)  
COMMENT Contact: Liew CC  
Brigham and Women's Hospital  
Harvard Medical School  
75 Francis St. Boston, MA 02115, USA  
Tel: 617328915  
Fax: 6179750995  
Email: cliew@rics.bwh.harvard.edu  
PCR Primers  
FORWARD: 5' GCCAGCTGGAATTAACCTCACTAAAGG 3'  
REVERSE: 5' CCGATGAATTTGATACGACTCACTAATAGG 3'  
Seq primer: 5' GAAATTAACCTCACTAAAGG 3'











Db 179 CAGATGGAGGCTGATACGATGATGATGTTGTTTCATCCGCAATTCCTGAGCTCA 238  
Qy 136 TAAACCCCTCACCCTGTTTACGATGATGTTTATAGGTTAGGCTCAGA 195  
Db 229 TAATCCCATAGCCCTTATACATGTTTGTATCATGTTGGGAGTAGTGCTCAGG 298  
Qy 196 AGCA 199  
Db 299 AGCA 302

RESULT 6  
LOCUS AL038727 617 bp mRNA linear EST 04-SEP-2003  
DEFINITION DKFZP566L1846\_r1 566 (synonym: hfk2) Homo sapiens cDNA clone  
AL038727  
ACCESSION AL038727.1 GI:5407877  
VERSION EST.  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 617)  
Ottewald, B., Obermaier, B., Mewes, H.W., Gassenhuber, J. and  
Wiemann, S.  
EST (Ottewald, et al.)  
TITLE Unpublished (1999)  
JOURNAL  
COMMENT Contact: MIPS

MIPS  
Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany  
This is the 5' sequence of the clone insert  
clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
Research Center (DKFZ), Email: s.wiemann@dkfz-heidelberg.de;  
sequenced by Medicinix (Martinsried/Germany) within the cDNA  
sequencing consortium of the German Genome Project. SI sequence  
also available.  
This clone (DKFZP566L1846) is available at the RZPD in Berlin.  
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
Berlin-Charlottenburg, GERMANY, Email: clone@rzpd.de.

FEATURES  
Location/Qualifiers

1..617  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="DKFZP566L1846"  
/tissue\_type="Kidney"  
/dev\_stage="fetal"  
/lab\_host="X1-2blue"  
/clone\_1lb="566 (synonym: hfk2)"  
/note="Vector: pAMP1, Site\_1: NotI, Site\_2: SalI"

ORIGIN

Query Match 23.4%; Score 70.4; DB 9; Length 617;  
Best Local Similarity 83.6%; Pred. No. 2, 1e-08;  
Matches 92; Conservative 0; Mismatches 16; Indels 2; Gaps 1;

Qy 188 GCGTCAGAGCAAACTCTCTCTCAGCTTCCAGCCCTCTCTCTCTGACACTCAT 247  
Db 342 GCCCAGAGCAAAAGTTCTCTCTGACCTTCTCCGCTCTCTCTCTGCTTTTCA 401  
Qy 248 TCTTCCCTGAGGC--CACATGAACATGAAATCTCTTCCACAGGCGG 295  
Db 402 TCTCCCAAGGCTACCATGAACTAGAACTCCCTCTTCCCAAGGCGAG 451

RESULT 7  
LOCUS BM983394/c 651 bp mRNA linear EST 20-FEB-2003  
DEFINITION UI-CF-DUI-aba-1-17-0-UI.s1 UI-CF-DUI Homo sapiens cDNA clone  
UI-CF-DUI-aba-1-17-0-UI 3', mRNA sequence.  
ACCESSION BM983394  
VERSION BM983394.1 GI:19609061

KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 651)  
Bonaldo, M.F., Lennon, G. and Soares, M.B.  
Normalization and subtraction: two approaches to facilitate gene  
discovery  
Genome Res. 6 (9), 791-806 (1996)  
JOURNAL 9704447  
MEDLINE 8889548  
PUBMED  
COMMENT Contact: McCray, PB  
McCray Lab  
University of Iowa  
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA  
Tel: 319 356 4866  
Fax: 319 356 7171  
Email: paul.mccray@uiowa.edu  
Tissue Procurement: Dr. M. J. Welsh, University of Iowa  
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research  
Genetics (www.resgen.com) or from Open Biosystems  
(www.openbiosystems.com).  
The following repetitive elements were found in this cDNA  
sequence: 23-564, >MER39B  
Seq primer: M13 FORWARD  
POLYA=yes.

FEATURES  
Location/Qualifiers

1..651  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-CF-DUI-aba-1-17-0-UI"  
/tissue\_type="Primary Lung Epithelial Cells"  
/dev\_stage="Adult"  
/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
/clone\_1lb="UI-CF-DUI"  
/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a  
modified polylinker; Site 1: EcoR I; Site 2: Not I;  
UI-CF-DUI is a normalized cDNA library containing the  
following tissue(s): Primary Lung Epithelial Cells The  
library was constructed according to Bonaldo, Lennon and  
Soares, Genome Research, 6:791-806, 1996. First strand  
cDNA synthesis was primed with an oligo-dT primer  
containing a Not I site. Double stranded cDNA was ligated  
to an EcoR I adaptor, digested with Not I, and cloned  
directionally into pT73-Pac vector. The oligonucleotide  
used to prime the synthesis of first-strand cDNA contains  
a library tag sequence that is located between the Not I  
site and the (dT)18 tail. The sequence tag for this  
library is GCGTCGAGGC.  
TAG\_TISSUE=Lung Epithelial Cells Tissue nos 359-368  
TAG\_LIB=UI-CF-DUI  
TAG\_SEQ=GCGTCGAGGC"

ORIGIN  
Query Match 23.4%; Score 70.4; DB 12; Length 651;  
Best Local Similarity 83.6%; Pred. No. 2, 1e-08;  
Matches 92; Conservative 0; Mismatches 16; Indels 2; Gaps 1;

Qy 188 GCGTCAGAGCAAACTCTCTCTCAGCTTCCAGCCCTCTCTCTGACACTCAT 247  
Db 538 GCCCAGAGCAAAAGTTCTCTCTGACCTTCTCCGCTCTCTCTGCTTTTCA 479  
Qy 248 TCTTCCCTGAGGC--CACATGAACATGAAATCTCTTCCACAGGCGG 295  
Db 478 TCTCCCAAGGCTACCATGAACTAGAACTCCCTCTTCCCAAGGCGAG 429

RESULT 8



BG741476  
 LOCUS 783 bp mRNA linear EST 15-MAY-2001  
 DEFINITION 602632271.F1 NCI CGAP\_Skn3 Homo sapiens cDNA clone IMAGE:4777553 5',  
 mRNA sequence.  
 ACCESSION BG741476  
 VERSION BG741476.1 GI:14052129  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NIH-MGC http://mgs.nci.nih.gov/.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 CONTACT: Robert Strausberg, Ph.D.  
 Email: cgapbs-remail.nih.gov  
 Tissue Procurement: James Cleaver, M.D.  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA  
 Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: LLM10631 row: n column: 18  
 High quality sequence stop: 760.  
 Location/Qualifiers  
 1. 783  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4777553"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /clone\_lib="NCI CGAP\_Skn3"  
 /note="Organ: skin; Vector: pCMV-SPORT6, Site\_1: NotI;  
 Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT.  
 Average insert size 1.5kb. Library constructed by Life  
 Technologies. Note: this is a NCI CGAP Library."  
 ORIGIN  
 Query Match 23.4%; Score 70.4; DB 12; Length 783;  
 Best Local Similarity 83.6%; Pred. No. 2.2e-08;  
 Matches 92; Conservative 0; Mismatches 16; Indels 2; Gaps 1;  
 QY 188 GCTCAGAGAAACTCTCTCTCACTTCTCCAGCCCTCCGTCTGCAACCTCAT 247  
 |||||  
 DB 243 GCCCAGAGAAAGTTCTCTCTGACCTTCTCCCTCTGTCTCTGCGCTTTTCAT 302  
 |||||  
 QY 248 TCTTCCCTGAGGC-CACATAGAACTAGATCTCTTCCACAGCGG 295  
 |||||  
 DB 303 TCTCCCAAGGCTACCCATAGAACTAGATCCCTTCCCAAGCGAG 352  
 |||||  
 RESULT 9  
 AM948684 303 bp mRNA linear EST 31-MAY-2000  
 LOCUS AM948684/c  
 DEFINITION QV0-FT0002-050500-230-f04 FT0002 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION AM948684  
 VERSION AM948684.1 GI:8126458  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 303)  
 Naeai M.A., da Silva M. Jr., Zago M.A., Bordin S., Costa F.F.,  
 Goldman G.H., Carvalho A.F., Matsukuma A., Bala G.S., Simpson D.H.,  
 Brunstein A., de Oliveira P.S., Bucher P., Jongeneel C.V.,  
 O'Hare M.J., Soares F., Brentani R.R., Reis L.F., de Souza S.J. and  
 Simpson A.J.  
 Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 MEDLINE 20202663  
 PUBMED 10737800  
 COMMENT Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the PAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?el=et2-qv0-FT0002-050  
 500-230-f04&3=2000-05-05&4=1)  
 Seq primer: puc 18 forward  
 High quality sequence stop: 303.  
 Location/Qualifiers  
 1. 303  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /dev\_stage="Adult"  
 /clone\_lib="FT0002"  
 /note="Organ: prostate tumor; Vector: puc18; Site\_1: SmaI;  
 Site\_2: SmaI; A mini-library was made by cloning products  
 derived from ORBATES PCR (U.S. Letters Patent application  
 No. 196,716 - Ludwig Institute for Cancer Research)  
 profiles into the puc 18 vector. Reverse transcription of  
 tissue mRNA and cDNA amplification were performed under  
 low stringency conditions."  
 ORIGIN  
 Query Match 22.9%; Score 68.8; DB 10; Length 303;  
 Best Local Similarity 82.7%; Pred. No. 5.2e-08;  
 Matches 91; Conservative 0; Mismatches 17; Indels 2; Gaps 1;  
 QY 188 GCTCAGAGAAACTCTCTCTCACTTCTCCAGCCCTCCGTCTGCAACCTCAT 247  
 |||||  
 DB 234 GCCCAGAGAAAGTTCTCTCTGACCTTCTCCCTCTGTCTCTGCGCTTTTCAT 175  
 |||||  
 QY 248 TCTTCCCTGAGGC-CACATAGAACTAGATCTCTTCCACAGCGG 295  
 |||||  
 DB 174 TCTCCCAAGGCTACCCATAGAACTAGATCCCTTCCCAAGCGAG 125  
 |||||  
 RESULT 10  
 BX362298 548 bp mRNA linear EST 05-MAY-2003  
 LOCUS BX362298  
 DEFINITION BX362298 Homo sapiens T CELLS (JURKAT CELL LINE) COT 10-NORMALIZED  
 Homo sapiens cDNA clone CS0DJ013YA03 5-PRIME, mRNA sequence.  
 ACCESSION BX362298  
 VERSION BX362298.1 GI:30368597  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 548)  
 Li W.B., Gruber C., Jessee J. and Polayes D.  
 Full-length cDNA libraries and normalization  
 Unpublished (2001)  
 CONTACT: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 5793.f for  
 more information about this cluster, see  
 http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CS0DJ013YA02P1c1cluster=5793.f. Contact :  
 Feng Liang Email : fliang@life-tech.com URL :  
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600







TITLE	REFERENCE AUTHORS
Use of BAC End Sequences from Library RCGI-11 for Sequence-Ready Map Building	Venter, J. C. Zhao, S., Adams, M. D., Niernan, W., Malek, J., de Jong, P. and
	Eukaryota; Metazoa; Chordata; Crustacea; Cnidaria; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Carnivora; Hominoidea; Homo.
	1 (bases 1 to 699)



JOURNAL  
COMMENT

Unpublished (1997)  
 Other GSSs: RPCI-11-205H11.TV  
 Contact: Shaying Zhao, William Niernman, Mark Adams  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: hbe@tigr.org  
 Clones are derived from the human BAC library RPCI-11. For BAC  
 library availability, please contact Pieter de Jong  
 (pieterdejong.med.buffalo.edu). Clones may be purchased from  
 BACPAC Resources (<http://bacpac.med.buffalo.edu/ordering>) or from  
 Research Genet cs ([info@resgen.com](mailto:info@resgen.com)). BAC end search page:  
[http://www.tigr.org/cdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/cdb/humgen/bac_end_search/bac_end_search.html).  
 Seq primer: T7  
 Class: BAC ends.  
 Location/Qualifiers

FEATURES  
source

1..699  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="GDB:7578514"  
 /db\_xref="taxon:9606"  
 /clone="RPCI-11-205H11"  
 /sex="Male"  
 /cell\_type="lymphocytes"  
 /clone\_lib="RPCI-11"  
 /note="Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
 RPCI11 Human Male BAC Library"

## ORIGIN

Query Match 22.5%; Score 67.8; DB 28; Length 699;  
 Best Local Similarity 73.0%; Pred. No. 1.1e-07;  
 Matches 119; Conservative 0; Mismatches 32; Indels 12; Gaps 2;  
 QY 87 TGTGATATATGATATATATAGTTTGTTCATAGTTCTGCTTATAAACCCTC 146  
 DB 441 TATGATTTATGATACATGATGTTTCATCAGAGCTCTGGCTCATTAATCTCTATA 382  
 QY 147 ACCCTGTACAGTCATTTGTATAAGTT-GGATGTTTAGGCTCAGAGCAAAACTC 205  
 DB 381 ACCCTGTATTTGCTTTTGTATAAGTTGGGCACTTAGGCTCAGAGGAGGCTC 322  
 QY 206 -----TCTCTCTACCTTCTCAGCCCTCTGTCTCT 237  
 DB 321 AGGAGCAGATCTCTGACCTTCTCGCTTCTTACCT 279

Search completed: March 13, 2004, 00:38:48  
 Job time : 962.585 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OW nucleic - nucleic search, using sw model

Run on: March 12, 2004, 12:27:59 ; Search time 947.629 Seconds

(without alignments)  
13767.261 Million cell updates/sec

Title: US-09-966-880A-10\_COPY\_6264\_6564

Perfect score: 301

Sequence: 1 gcatcaagaagacacttagta.....cttcacaagaagcgcgcaagaag 301

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_ba: \*  
2: gb\_hg: \*  
3: gb\_in: \*  
4: gb\_cm: \*  
5: gb\_ov: \*  
6: gb\_pac: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_sta: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vl: \*  
15: gb\_ba: \*  
16: em\_fun: \*  
17: em\_hum: \*  
18: em\_in: \*  
19: em\_mu: \*  
20: em\_cm: \*  
21: em\_or: \*  
22: em\_ov: \*  
23: em\_pat: \*  
24: em\_ph: \*  
25: em\_pl: \*  
26: em\_ro: \*  
27: em\_sta: \*  
28: em\_un: \*  
29: em\_vl: \*  
30: em\_hg\_hum: \*  
31: em\_hg\_inh: \*  
32: em\_hg\_other: \*  
33: em\_hg\_mus: \*  
34: em\_hg\_pin: \*  
35: em\_hg\_rod: \*  
36: em\_hg\_mam: \*  
37: em\_hg\_vrt: \*  
38: em\_sy: \*  
39: em\_hgo\_hum: \*  
40: em\_hgo\_mus: \*  
41: em\_hgo\_other: \*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	301	100.0	6564	6 BD016835	BD016835 Novel cyt
2	301	100.0	71132	9 AC002184	AC002184 Homo sapi
3	89.4	29.7	39395	9 AC004212	AC004212 Homo sapi
4	89.4	29.7	76225	9 AL773551	AL773551 Human DNA
5	89.4	29.7	112018	9 AB023050	AB023050 Homo sapi
6	89.4	29.7	171627	9 AL662797	AL662797 Human DNA
7	89.4	29.7	145617	9 AL662848	AL662848 Human DNA
8	89.4	29.7	200000	9 AP000512	AP000512 Homo sapi
9	87.8	29.2	300050	9 AB100084	AB100084 Pan trogl
10	83.8	27.8	481	9 HSDJ93K2	HSDJ93K2
11	83.8	27.8	113804	9 AC025610	AC025610 Homo sapi
12	83.8	27.4	143133	2 AC021655	AC021655 Homo sapi
13	82.6	27.4	163862	2 AC092911	AC092911 Homo sapi
14	82.6	27.4	165082	2 AC026559	AC026559 Homo sapi
15	82.6	27.4	154006	9 AC074044	AC074044 Homo sapi
16	82.6	27.4	229396	2 AC024108	AC024108 Homo sapi
17	82.6	27.1	123742	9 AC069431	AC069431 Homo sapi
18	81.6	27.1	152433	2 AC036127	AC036127 Homo sapi
19	81.6	27.0	123057	2 BX890549	BX890549 Dario rer
20	81.2	27.0	195117	9 AL353741	AL353741 Human DNA
21	81.2	26.8	59561	2 AC131265	AC131265 Homo sapi
22	80.6	26.8	173535	9 AC021316	AC021316 Homo sapi
23	80.6	26.8	173986	9 AC104759	AC104759 Homo sapi
24	80.6	26.3	98300	9 AC006236	AC006236 Homo sapi
25	79.2	26.1	176368	2 AC140859	AC140859 Homo sapi
26	78.6	26.1	185737	2 AC145616	AC145616 Homo sapi
27	78.6	26.1	197637	9 AC023141	AC023141 Homo sapi
28	78.6	26.1	207614	2 AC098800	AC098800 Homo sapi
29	78.6	26.0	171683	9 AC105179	AC105179 Homo sapi
30	77.8	25.8	140137	2 AC027611	AC027611 Homo sapi
31	77.8	25.8	151847	2 AC022792	AC022792 Homo sapi
32	77.8	25.8	165378	9 AC129805	AC129805 Homo sapi
33	77.6	25.8	147788	9 AP005360	AP005360 Homo sapi
34	77.6	25.8	152440	9 AC104233	AC104233 Homo sapi
35	77.6	25.8	162241	9 AC022909	AC022909 Homo sapi
36	77.6	25.8	184533	9 AC093592	AC093592 Homo sapi
37	77.6	25.8	192573	9 AC073344	AC073344 Homo sapi
38	77.4	25.7	158562	9 AL603888	AL603888 Human DNA
39	77.4	25.5	110000	2 AL929091_02	Continuation (3 of
40	76.8	25.5	110000	2 AL929091_03	Continuation (4 of
41	76.8	25.5	155889	2 AC025642	AC025642 Homo sapi
42	76.8	25.5	162347	9 AL356108	AL356108 Human DNA
43	76.4	25.4	216451	9 AC098818	AC098818 Homo sapi
44	76.4	25.4	219207	2 AC009685	AC009685 Homo sapi
45	76.4	25.4	219207	2 AC009685	AC009685 Homo sapi

#### ALIGNMENTS

RESULT 1  
BD016835  
LOCUS BD016835  
DEFINITION Novel cytidine deaminase.  
ACCESSION BD016835  
VERSION BD016835.1 GI:22558011  
KEYWORDS UP 2001245669-A/8.  
SOURCE  
ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 6564)  
AUTHORS Honjo, T. and Muramatsu, M.  
TITLE Novel cytidine deaminase  
JOURNAL Patent: JP 2001245669-A 8 11-SEP-2001;



COMMENT  
JAPAN TOBACCO INC. TASUKU HONJO  
OS Homo sapiens (human)  
PN JP 2001245669-A/8  
PD 11-SEP-2001  
PF 28-MAR-2000 JP 2000092981  
PI TASUKU HONJO, MASAMICHI MIMAMATSU  
PC C12N15/09, A61K39/395, A61K39/395, A61P1/00, A61P11/06, A61P13/12,  
PC A61P17/00,  
PC A61P27/02, A61P27/16, A61P37/02, A61P37/08, C07K16/18, C12N1/19, PC  
C12N1/21,  
PC C12N5/10, C12N9/78, C12P21/02, C12P21/08 // (C12N1/21, C12R1/19), PC  
(C12N5/10, C12R1/91), C12N15/00, C12N5/00, (C12N5/00, C12R1/91) CC  
FH Key Location/Qualifiers.  
FEATURES  
source  
1. .6564  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Query Match 100.0%; Score 301; DB 6; Length 6564;  
Best Local Similarity 100.0%; Pred. No. 1e-71;  
Matches 301; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GCATCAAGGACACTAGTAAAGGAGAAAGCAAGCAAGGAGGAGAAACAC 60  
DB 6264 GCATCAAGGACACTAGTAAAGGAGAAAGCAAGCAAGGAGGAGAAACAC 6323  
QY 61 TTGCAAACTATATCTGATAGGGTTGGATTTATGATATATATATAGTTTTC 120  
DB 6324 TTGCAAACTATATCTGATAGGGTTGGATTTATGATATATATATAGTTTTC 6383  
QY 121 ATAGTTCCTGGCTTATTAACCCCTCACCCCTTACAGTCTTGTATAGTTGAT 180  
DB 6384 ATAGTTCCTGGCTTATTAACCCCTCACCCCTTACAGTCTTGTATAGTTGAT 6443  
QY 181 GGTTAGGCTCTGAGAACAACTCTCTCTCACCTTCTCCAGCCCTCTCTCTGAC 240  
DB 6444 GGTTAGGCTCTGAGAACAACTCTCTCTCACCTTCTCCAGCCCTCTCTCTGAC 6503  
QY 241 ACCCTATTTCTCTCTGAGGCGACATAGAACTGAACTCTCTCTTCCACAGGCGTCAA 300  
DB 6504 ACCCTATTTCTCTCTGAGGCGACATAGAACTGAACTCTCTCTTCCACAGGCGTCAA 6563  
QY 301 G 301  
DB 6564 G 6564  
RESULT 2  
LOCUS AC092184 71132 bp DNA linear PRI 12-UTN-2002  
DEFINITION Homo sapiens 12 BAC RP11-438L7 (Roswell Park Cancer Institute Human  
AC092184 AC013443  
ACCESSION AC092184.7 GI:21206067  
VERSION  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 71132)  
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-ouman, F.R., Allen, C.,  
Alsbrooks, S.L., Amaralunge, H.C., Are, J.R., Ayale, M., Banks, T.,  
Barbaria, J., Benton, J., Bimaga, K., Blankenburg, K., Bonin, D.,  
Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,  
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,  
Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,  
Chen, G., Chen, R., Chen, Z., Chiu, D., Chowdhury, I., Christopoulos, C.,  
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorn, R., David, R.,  
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,  
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,  
Doutwalte, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Worley, K.C.  
Direct Submission  
Submitted (25-MAY-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 71132)  
Worley, K.C.  
Direct Submission  
Submitted (18-MAY-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
4 (bases 1 to 71132)  
Worley, K.C.  
Direct Submission  
Submitted (25-MAY-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
5 (bases 1 to 71132)  
Worley, K.C.  
Direct Submission  
Submitted (12-UTN-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On May 25, 2002 this sequence version replaced gi:20901754.  
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email  
gc-help@bcm.tmc.edu  
CLONE LENGTH: This sequence does not necessarily represent the  
entire insert of this clone. Overlapping regions of clones are only  
sequenced and submitted once, so the sequence for the remainder of  
the insert may be found in the record for the adjacent clones.  
Overlapping clones are noted at the beginning and end of the  
Features listing.  
ANNOTATION OF FEATURES:  
STS are identified using ePCR (Genome Res. 7:541-550) searches  
of a local database that includes entries from dbSTS, GDB, and  
local mapping efforts.  
Repeats are identified using RepeatMasker (A. Smit and P. Green,  
Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Emerling, S.,  
Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P.,  
Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N.,  
Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,  
Hamilton, C., Han, J., Harris, C., Harris, K., Hart, M., Havlak, P.,  
Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M.,  
Holloway, C., Hollins, B., Homsi, F., Howard, S., Huber, J., Hulik, S.,  
Hume, J., Ioshikhes, I., Jackson, L.E., Jacobson, B., Jia, Y.,  
Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U.,  
Kang, L., Korvah, J., Kovar, C., Kratovic, J., Kureishi, A., Landry, N.,  
Leal, B., Lee, E., Lewis, L.C., Lewis, J., Li, J., Li, Z., Licharge, O.,  
Lieu, C., Liu, J., Liu, W., Louised, H., Lozano, R.J., Lu, X.,  
Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapa, P.,  
Marandel, I., Martin, R., Martindale, A., Martinez, E., Massey, E.,  
Mawhinney, E., McLeod, M.P., Meador, M., Mei, G., Merscher, S.,  
Metzker, M., Miller, A., Miner, G., Miner, Z., Mitchell, T.,  
Mohabbat, K., Montgomery, K.T., Morgan, M., Morris, S., Moser, M.,  
Neal, D., Nelson, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okunou, G.,  
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, D., Perez, L.,  
Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,  
Rivers, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savery, G.,  
Scherer, S., Scott, G., Shen, H., Shim, C., Shochet, N., Sisson, I.,  
Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,  
Sutton, A., Swatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
Umani, K., Vasquez, L., Vera, V., Villalón, D., Vinton, R., Wang, Q.,  
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S.,  
Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K.,  
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Kucherlapati, R.,  
Weinstock, G., and Gibbs, R.  
Direct Submission  
Unpublished  
2 (bases 1 to 71132)  
Worley, K.C.  
Direct Submission  
Submitted (25-JUN-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 71132)  
Worley, K.C.  
Direct Submission  
Submitted (18-MAY-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
4 (bases 1 to 71132)  
Worley, K.C.  
Direct Submission  
Submitted (25-MAY-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
5 (bases 1 to 71132)  
Worley, K.C.  
Direct Submission  
Submitted (12-UTN-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On May 25, 2002 this sequence version replaced gi:20901754.  
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email  
gc-help@bcm.tmc.edu



unpublished.) for Human and Mouse sequences. Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING LEAD COVERAGE: sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES. This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:  
[http://gc.bom.tmc.edu/quality\\_info/genbank\\_annotation.html](http://gc.bom.tmc.edu/quality_info/genbank_annotation.html).

## QUALSTAT-REPORT

FEATURES	Location/Qualifiers
source	1. .71132

```

repeat_region complement(6711, .6617)
/rpt_family="MIR"
repeat_region complement(7033, .7327)
/rpt_family="AluXc"
repeat_region 7607, .7745
/rpt_family="FLAM_A"
repeat_region 7980, .8047
/rpt_family="MIR"
repeat_region complement(8050, .8549)
/rpt_family="LTR47A"
repeat_region 8567, .8698
/rpt_family="MIR"
repeat_region complement(8815, .9101)
/rpt_family="AluBg"
repeat_region complement(11227, .12175)
/rpt_family="LTR5"
STS
12625, .12900
/standard_name="87432"
repeat_region complement(13547, .13664)
/rpt_family="L1MB5"
repeat_region complement(13665, .13979)
/rpt_family="AluY"
repeat_region complement(13980, .14193)
/rpt_family="L1MB5"
repeat_region 14622, .14924
/rpt_family="AluY"
repeat_region complement(15813, .17136)
/rpt_family="L1PA8"
repeat_region complement(17272, .17395)
/rpt_family="FLAM_C"
repeat_region 17426, .17575
/rpt_family="MIR"

```

Query Match	Score	DB	Length
100.0%	301	9	71132

```
Matches 301; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy	1	GCATCAAAAGCACTTAGTAAAGAACTGAAAAAGACACACCACAGAAAGTGGAGAAAAAC	60
Dd	46022	GCATCAAAAGCACCTTAGTAAAGAAAGTAAAAAGACACACCACAGAAAGTGGAGAAAAAC	46081
Qy	61	TTGCAATCATATATCTGATTAAGGTTGTGATATATAGTATATATATAGTTTGTGC	120
Dd	46082	TTGCAATCATATATCTGATTAAGGTTGTGATATATAGTATATATATAGTTTGTGC	46141
Qy	121	ATAGTCTTGCTGATTATTAACCCCTGCACCCTGTATACATCATTTGTATTAAGTTGAT	180
Dd	46142	ATAGTCTTGCTGATTATTAACCCCTGCACCCTGTATACATCATTTGTATTAAGTTGAT	46201
Qy	181	GGTTTAGGCTTCAGAAACAAACTCTCTCTCTCACTTTCAGACCTCTCTGTCTTGCC	240
Dd	46202	GGTTTAGGCTTCAGAAACAAACTCTCTCTCTCACTTTCAGACCTCTCTGTCTTGCC	46261
Qy	241	ACCTCATTTCTCCCTTAGGCGCATATGAACATAGATCTCTTTCCACAGGCGGTCAAA	300
Dd	46262	ACCTCATTTCTCCCTTAGGCGCATATGAACATAGATCTCTTTCCACAGGCGGTCAAA	46321
Qy	301	G	301
Dd	46322	G	46322

RESULT 3	AC004212	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
AC004212	39395 bp	DNA	linear	PRI 17-JUL-1998				
AC004212	Homo sapiens clone UMG0C:y67c126	from 6p21,	complete sequence.					
AC004212								
AC004212.1	GI:3327404							
HTG.								
Source	Homo sapiens (human)							
Organism	Homo sapiens							
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;							
	Mammalia; Eutheria; Primates; Catarrhini; Homiindae; Homo.							



REFERENCE 1 (bases 1 to 39395)  
AUTHORS Janer,M.M., Guillaudoux,T., Vu,Q., Kutyavin,T., Harter,H. and  
Geraghty,D.E.  
TITLE Large scale sequence analysis of the human MHC class I region  
JOURNAL Unpublished (1998)  
REMARK Fred Hutchinson Cancer Research Center  
The Clinical Research Division  
1100 Fairview Ave. N., P.O. Box 19024  
Seattle, WA 98109-1024

REFERENCE 2 (bases 1 to 39395)  
AUTHORS Geraghty,D.E. and Olson,M.V.  
TITLE Direct Submission  
JOURNAL Submitted (23-FEB-1998) Human Genome Center, University of  
Washington, Box 352145, Seattle, WA 98195, USA  
3 (bases 1 to 39395)  
Geraghty,D.E. and Olson,M.V.  
Direct Submission  
Submitted (17-JUL-1998) Human Genome Center, University of  
Washington, Box 352145, Seattle, WA 98195, USA  
Contact: Daniel E. Geraghty (geraghty@fhcc.org)  
On Jul 17, 1998 this sequence version replaced gi:2905848.  
Overlapping Sequences:  
5': UMGC:Y67C092  
3': UMGC:Y54C283

REMARK  
COMMENT On Jul 17, 1998 this sequence version replaced gi:2905848.  
Overlapping Sequences:  
5': UMGC:Y67C092  
3': UMGC:Y54C283

#### Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

Double stranded (DS) coverage: 89.8%  
DS or two chemistry coverage: 100.0%  
Single stranded regions: 0

#### Sequence Validation:

This sequence has been validated by Multiple Complete Digest Mapping. Comparison of the experimentally derived map digest fragments with sequence-predicted fragments is given below. Small fragments below a variable cutoff (approximately 400-600bp) are not mapped and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragment groups are separated by dashed lines.

Map	Seq	Map	Seq	Map	Seq
13333.40	12934.00	4775.80	4608.00	5063.92	4866.00
10170.57	9750.00	8191.44	7867.00	560.75	556.00
3724.57	3591.00	11693.00	11114.00	2711.00	2644.00
5055.83	4873.00			13108.62	12618.00
3403.86	3333.00			876.14	866.00
1302.20	1294.00			10536.25	10080.00

#### FEATURES

##### source

1. 39395  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/map="6p21"  
/clone="CGM1:A146A3"  
/sub\_clone="UMGC:Y67C126"  
/cell\_line="CGM1"

repeat\_region /clone\_lib="Wash U YAC library"  
repeat\_region complement(11..498)  
repeat\_region /rpt\_family="Alu"  
repeat\_region complement(11793..2035)  
repeat\_region /rpt\_family="Alu"  
repeat\_region complement(2072..2364)  
repeat\_region /rpt\_family="Alu"  
repeat\_region 2902..3188  
repeat\_region /rpt\_family="Alu"  
repeat\_region complement(3297..3710)  
repeat\_region /rpt\_family="MLT2B2"  
repeat\_region complement(3872..4138)  
repeat\_region /rpt\_family="Alu"  
repeat\_region complement(5128..5284)  
repeat\_region /rpt\_family="Alu"  
repeat\_region complement(9979..10294)  
repeat\_region /rpt\_family="Alu"  
repeat\_region 10379..10492  
repeat\_region /rpt\_family="MIR"  
repeat\_region 10872..11146  
repeat\_region /rpt\_family="MIR"  
repeat\_region complement(11926..12095)  
repeat\_region /rpt\_family="Alu"  
repeat\_region complement(12242..12529)  
repeat\_region /rpt\_family="Alu"  
repeat\_region 12823..12874  
repeat\_region /rpt\_family="Alu"  
repeat\_region complement(13004..13295)  
repeat\_region /rpt\_family="Alu"  
repeat\_region 14863..15437  
repeat\_region /rpt\_family="Alu"  
repeat\_region 16992..17282  
repeat\_region /rpt\_family="Alu"  
repeat\_region complement(17924..18229)  
repeat\_region /rpt\_family="Alu"  
repeat\_region 18513..18732  
repeat\_region /rpt\_family="Alu"  
repeat\_region 19271..19498  
repeat\_region /rpt\_family="Alu"  
repeat\_region 20131..21084  
repeat\_region /rpt\_family="L1R12"  
repeat\_region 21145..21195  
repeat\_region /rpt\_family="Alu"  
repeat\_region 21196..21470  
repeat\_region /rpt\_family="Alu"  
repeat\_region 22542..22722  
repeat\_region /rpt\_family="Alu"  
repeat\_region 23189..23488  
repeat\_region /rpt\_family="Alu"  
repeat\_region 23542..23831  
repeat\_region /rpt\_family="Alu"  
repeat\_region complement(24444..25026)  
repeat\_region /rpt\_family="Alu"  
repeat\_region complement(26092..26389)  
repeat\_region /rpt\_family="Alu"  
repeat\_region complement(26824..27046)  
repeat\_region /rpt\_family="Alu"  
repeat\_region complement(27109..27405)  
repeat\_region /rpt\_family="Alu"  
repeat\_region 27692..27976  
repeat\_region /rpt\_family="Alu"  
repeat\_region 28990..29615  
repeat\_region /rpt\_family="L1R12"  
repeat\_region complement(29635..29824)  
repeat\_region /rpt\_family="MER39"  
repeat\_region complement(29635..29751)  
repeat\_region /rpt\_family="MER21"  
repeat\_region complement(29853..30134)  
repeat\_region /rpt\_family="Alu"  
repeat\_region complement(30162..30221)  
repeat\_region /rpt\_family="MER39"  
repeat\_region 30431..30708  
repeat\_region /rpt\_family="Alu"



```

repeat_region complement (30836, .31706)
repeat_region /rpt_family="MER21"
repeat_region complement (30895, .31195)
repeat_region /rpt_family="MER39"
repeat_region complement (32270, .32547)
repeat_region /rpt_family="Alu"
repeat_region complement (32599, .33195)
repeat_region /rpt_family="Alu"
repeat_region complement (33661, .33946)
repeat_region /rpt_family="Alu"
repeat_region 34848, .35115
repeat_region /rpt_family="Alu"
repeat_region 35730, .35916
repeat_region /rpt_family="Alu"
repeat_region 36293, .36408
repeat_region /rpt_family="Alu"
repeat_region 37656, .37902
repeat_region /rpt_family="MER42"
variation 38541
variation /note="high quality variation versus 5' overlapping clone"
repeat_region 38547, .38701
repeat_region /rpt_family="Alu"
repeat_region 38843, .39083
repeat_region /rpt_family="Alu"
variation 39213
variation /note="high quality variation versus 3' overlapping clone"

```

## ORIGIN

```

Query Match 29.7%; Score 89.4; DB 9; Length 39395;
Best Local Similarity 73.7%; Pred. No. 6.9e-14;
Matches 132; Conservative 0; Mismatches 36; Indels 11; Gaps 1;

```

```

Qy 87 TGTGATATATGATATATATAGCTTTTGTCCATGTTCTGGCTTAAACCCCTC 146
Db 30811 TGTCTATATATATATATATAGCTTTTGTCCATGTTCTGGCTTAAATCCCAT 30870
Qy 147 ACCCTGTTACAGCATTTGTTATAGGTTGATGTTAGGCTCGAAGCAAACTC- 205
Db 30871 GCCCTGTACCATCTGTTGTTATAGTGGGCTTTAAGCTCGAAGCAAGTCTCA 30930
Qy 206 -----TCTCTACACCTTCTCCAGCCCTCTGCTCTGCTGCAACCTCATTTCC 254
Db 30931 ATAAAGGAATCTCTGACCTTCTCGCCCTCATTTCACTGCTCTTATCTCTCC 30989

```

```

RESULT 4
AL773551/c 76225 bp DNA linear PRI 09-AUG-2002
LOCUS Human DNA sequence from clone Xkbac-230F21 on chromosome 6,
DEFINITION complete sequence.
ACCESSION AL773551
VERSION AL773551.4 GI:22204642
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

```

```

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 76225)
Almeida, J.
JOURNAL Direct Submission
Submitted (01-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk; Clone requests: clonerequest@sanger.ac.uk
On Aug 11, 2002 this sequence version replaced gi:21621738.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk

```

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the

variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M3 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs constructed by the MHC HaploType Consortium and collaborators. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6/MHC> Xkbac-230F21 is from a DNA-arts QBL human bac library VECTOR: pBelobac11.

## FEATURES

```

source Location/Qualifiers
1..76225
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="6"
/clone="Xkbac-230F21"
/clone_id="DNA-arts-BAC.1-QBL.1"

```

## ORIGIN

```

Query Match 29.7%; Score 89.4; DB 9; Length 76225;
Best Local Similarity 73.7%; Pred. No. 6.7e-14;
Matches 132; Conservative 0; Mismatches 36; Indels 11; Gaps 1;

```

```

Qy 87 TGTGATATATGATATATATAGCTTTTGTCCATGTTCTGGCTTAAACCCCTC 146
Db 59998 TGTGCTATATATATATATATAGCTTTTGTCCATGTTCTGGCTTAAATCCCAT 59939
Qy 147 ACCCTGTTACAGCATTTGTTATAGGTTGATGTTAGGCTCGAAGCAAACTC- 205
Db 59938 GCCCTGTACCATCTGTTGTTATAGTGGGCTTTAAGCTCGAAGCAAGTCTCA 59879
Qy 206 -----TCTCTACACCTTCTCCAGCCCTCTGCTCTGCTGCAACCTCATTTCC 254
Db 59878 ATAAAGGAATCTCTGACCTTCTCGCCCTCATTTCACTGCTCTTATCTCTCC 59820

```

```

RESULT 5
AB023050 112018 bp DNA linear PRI 22-NOV-2000
LOCUS Homo sapiens genomic DNA, chromosome 6p21.3, HLA class I region,
DEFINITION clone:832F2, complete sequence.
ACCESSION AB023050
VERSION AB023050.1 GI:5672605
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (sites)
REFERENCE
AUTHORS Shihina, T., Tamiya, G., Oka, A., Takishima, N., Yamagata, T.,
Kikkawa, E., Iwata, K., Tomizawa, M., Okuaki, N., Kuwano, Y.,
Watanabe, K., Fukuzumi, Y., Itakura, S., Sugawara, C., Ono, A.,
Yamazaki, M., Taahiro, H., Ando, A., Iemura, T., Soeda, E., Kimura, M.,
Baham, S. and Inoko, H.
Molecular dynamics of MHC genes is unraveled by sequence analysis of
the 1,796,938-bp HLA class I region
Proc. Natl. Acad. Sci. U.S.A. 96 (23), 13282-13287 (1999)
JOURNAL MEDLINE
PUBMED 20027539
10577312
2 (sites)
REFERENCE
AUTHORS Keicho, N., Ohashi, J., Tamiya, G., Nakata, K., Taguchi, Y., Azuma, A.,

```



TITLE  
Ohishi,N., Emi,M., Park,M.H., Imoko,H., Tokunaga,K. and Kudoh,S.  
Fine localization of a major disease-susceptibility locus for  
diffuse panbronchiolitis  
Am. J. Hum. Genet. 66 (2), 501-507 (2000)

JOURNAL  
MEDLINE  
PUBMED  
20143375  
10677310  
3 (bases 1 to 112018)  
AUTHORS  
Shima,T. and Takishima,N.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (29-JUN-1999) Takashi Shima, Tokai University School of  
Medicine, Department of Molecular Life Science 2; Bohseida,  
Isehara, Kanagawa 259-1193, Japan  
(E-mail: tshima@is.icc.u-tokai.ac.jp, Tel:81-463-93-1121,  
Fax:81-463-94-8884)

FEATURES  
SOURCE  
Location/Qualifiers  
1. 112018  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/map="6p21.3"  
/clone="832P2"

## ORIGIN

Query Match 29.7%; Score 89.4; DB 9; Length 112018;  
Best Local Similarity 73.7%; Pred. No. 6,5e-14;  
Matches 132; Conservative 0; Mismatches 36; Indels 11; Gaps 1;

QY 87 TGTGATATTATGATATATATAGAGTTTGTTCATAGTCTGGCTTAAACCCCTC 146  
DB 111659 TGTGCTATTATATTATATATAGTTTATGTCATGCTTCGGCTTAATTTCCCAT 111718  
QY 147 ACCCTTTACAGCATTTGTTATAGTTGATGTTTGGCTTGAAGCAAAATC- 205  
DB 111719 GCCCTTTACATCTGTTGTTATGTTGGGCTTTAAGCTTGAAGCAAGAGTCTCA 111778  
QY 206 -----TCTCTTCACCTTCTCCAGCCCTCTCTCTCTGCAACCTCATCTTCCC 254  
DB 111779 ATAAACGAATCTCTGACCTTCTCTGCTCCCTCATTTCACTGCTTATCTCTCC 111837

RESULT 6  
AL662797/c 171627 bp DNA linear PRI 25-APR-2002  
LOCUS Human DNA sequence from clone Xbac-252P9 on chromosome 6, complete  
SEQUENCE.  
ACCESSION AL662797  
VERSION AL662797.7 GI:20338494  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 171627)  
Almeida,J.  
Direct Submission  
Submitted (25-APR-2002) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
On Apr 29, 2002 this sequence version replaced gi:20338227.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following

abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em, EMBL; Sw,  
SWISSPROT; Tr, TREMBL; Wp, WORMPEP; information on the WORMPEP  
database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence  
was generated from part of bacterial clone contigs constructed by  
the MHC HaploTYPE Consortium and collaborators. Further information  
can be found at  
http://www.sanger.ac.uk/HGP/Chre/MHC  
Xbac-252P9 is from a CHORI-501 human bac - PGF cell line library  
VECTOR: pPARAC2.1.

## FEATURES

SOURCE  
Location/Qualifiers  
1. 171627  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/clone="Xbac-252P9"  
/clone\_id="CHORI-501"

## ORIGIN

Query Match 29.7%; Score 89.4; DB 9; Length 171627;  
Best Local Similarity 73.7%; Pred. No. 6,4e-14;  
Matches 132; Conservative 0; Mismatches 36; Indels 11; Gaps 1;

QY 87 TGTGATATTATGATATATATAGAGTTTGTTCATAGTCTGGCTTAAACCCCTC 146  
DB 128619 TGTGCTATTATATTATATATAGTTTATGTCATGCTTCGGCTTAAATTTCCCAT 128560  
QY 147 ACCCTTTACAGCATTTGTTATAGTTGATGTTTGGCTTGAAGCAAAATC- 205  
DB 128559 GCCCTTTACATCTGTTGTTATGTTGGGCTTTAAGCTTGAAGCAAGAGTCTCA 128500  
QY 206 -----TCTCTTCACCTTCTCCAGCCCTCTCTCTCTGCAACCTCATCTTCCC 254  
DB 128499 ATAAACGAATCTCTGACCTTCTCTGCTCCCTCATTTCACTGCTTATCTCTCC 128441

RESULT 7  
AL662848/c 185617 bp DNA linear PRI 24-APR-2002  
LOCUS Human DNA sequence from clone Xbac-111D4 on chromosome 6, complete  
SEQUENCE.  
ACCESSION AL662848  
VERSION AL662848.6 GI:19031747  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 185617)  
Tracey,A.  
Direct Submission  
Submitted (23-APR-2002) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
On Feb 28, 2002 this sequence version replaced gi:1947645.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em, EMBL; Sw,  
SWISSPROT; Tr, TREMBL; Wp, WORMPEP; information on the WORMPEP



database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) Xxbac-111D4 is from a CHORI-502 human bac - COX cell line library VECTOR:

This sequence was generated from part of bacterial clone contigs constructed by the MHC Haplotype Consortium and collaborators. Further information can be found at <http://www.sanger.ac.uk/HGP/chr6/MHC>.

## FEATURES

```
source
1. 185617
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="axon:9606"
/chromosome="6"
/clone=XXDAC-11D4"
/clone_11b="CHORI-502"
```

ORIGIN

Query Match	29.7%;	Score 89.4;	DB 9;	Length 185617;
Best Local Similarity	73.7%;	Pred. No. 6.4e-14;		
Matches 132; Conservative	0;	Mismatches 36;	Indels 11;	Gaps 1;

Oy	87	TGTGATTTAAGTATATATATAGGTTTGTGCCATAGTCTCGGCTATAAACCCCTC	146
Db	113698	TGTGCTATTATATATATATAGGTTTGTGCCATAGTCTCGGCTATAATCCCAT	1136939

Oy	147	ACCCTTGTACAGTCATTTGGTATTAAGGTGTGATGTATTAAGCCCTCAGAGCAAAATC-	205
Db	113338	GCCTTTTACCATCTGTGTGTTGAATGTGTGGTCTTTAAGCCTCAGAGCAGGTCTCA	113579

QY	206	-----TCTCTTCACCGTCTCTCAGGCCCTCCCTGCTCTGGACCTCAATCTTCCC	254
Db	113578	ATPAAACGAAATCTCTCTCACTTCTCTCGCCCTCAATTCATCTGCTCTTATCTCTCC	113582

RESULT 8

LOCUS	AP000512	200000 bp	DNA	linear	PRI 22-AUG-2000
DEFINITION	Homo sapiens genomic DNA, chromosome 6p21.3, HLA Class I region, section 11/20.				

SOURCE ORGANISM	Homo sapiens (human)	Homo sapiens
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

REFERENCE  
AUTHORS 1  
TITLES Shima,S., Tamtja,G., Oka,A. and Inoko,H.  
JOURNAL Homo sapiens 2,123,817bp genomic DNA of 6p21.3 H1A class I region  
PUBLISHED ONLY IN DATABASE (1999)  
2 (bases 1 to 200000)  
AUTHORS Hirakawa,M., Yamaguchi,H., Imai,K. and Shimada,J.  
TITLES Direct Submision  
JOURNAL Submitted (21-SEP-1999) Mika Hirakawa, Japan Science and Technology

**COMMENT** This sequence is conducted by Tokai University as a JST sequencing Team.

Principal investigator: Hidetoshi Inoko Ph.D.  
Phone:+81-463-93-1121, Fax:+81-463-94-8884,  
The sequence is submitted by Human Genome Sequencing in ALIS  
project of JST

Japan science and technology corporation (jst)  
5-3, Yodobancyo, Chiyoda-ku, Tokyo, 102-0081 Japan  
For further information about this sequences, please visit our  
sequence archive web site (<http://www.aits.tokyo.iist.go.jp/HGS/top.html>) or send email to [webmaster@www.aits.tokyo.iist.go.jp](mailto:webmaster@www.aits.tokyo.iist.go.jp).

## FEATURES

```
source 1. .200000
/organism="Homo sapiens"
```

```

/mol type="genomic DNA"
/db xref="taxon:9606"
/chirosome="6"
/map="6p21.3"
97634..98938
/gene="PRG1"
97634..97872
/gene="PRG1"
/number=1
join(97653..97872,97985..98245)

```

```
exon      99978..100034  
          /gene="FGF1"  
          /number=2  
CDS       5161(99992)..100034,100314..100389,100477..100567  
          100848..100991,101383..101502,101590..101655,  
          101871..102023,111089..111270,111462..111557,  
          111624..111761,112003..112167,111404..114073
```

/tr/ana/na="MEFPCQENEMAVVSGSCRPWMAVAGAFVFLPQCIORJARS  
 TULINAKSKVYTRGVPIISVTGI/AVKTIQONKXEMLAAGAFKTEPAEIALIAL  
 TIEGHORAIMAMTWEETKYDROKSPSSVREKASBPLNMGISVYSUTJIKOHDHOPD  
 LHSIGKQAPRHOIXDAPRIGEPADRGATPEAKRQESYAOVLSITENAKRQPEY  
 KKAQYITVETRAADADAYOUQYAKTKOQIBSORQVQVVERAQQAYANOBETIRBP  
 KESVQV  
 AEBQAKKAAFPYQLEPAQDMLEKUPQVAEETISEGLTSANNTIVSGSGSGMGAA  
 VTSEVSDIITRIPESVERLTGTSISQVNHKPLRTA"  
 100314 .100389

```

exon      /gene="PLOT1"
           /number=3
           100477..100567
exon      /gene="PLOT1"
           /number=4
           100848..100991

```

```

/number=5
101383..101502
/gene="FLOT1"
/number=6
101440..101648
STS

```

```
/note="stSG4264  
of PCR primers.  
/_db_xref="GDB:4  
101590. .101685  
/_gene="FIOT1"  
exon
```

```

exon      /number=1
          101871..102023
          /gene="FLOT1"
          /number=8
exon      111089..111270

```

```

/number=9
111462. 111507
exon

```



exon /gene="PLOT1"  
/number=10  
111624, 111761  
/gene="PLOT1"  
/number=11  
112003, 112167  
/gene="PLOT1"  
/number=12  
114044, 114460  
/gene="PLOT1"  
/number=13  
complement (114197, 114328)  
/standard.name="D6s2115"  
/note="SHGC-13451; The location is between each flanking  
site of PCR primers."  
/db\_xref="GDB:738998"  
complement (114300, 114430)  
/note="WI-18721; The location is between each flanking site  
of PCR primers."  
/db\_xref="GDB:4574312"  
117236, 117362  
/note="SGC3496; The location is between each flanking site  
of PCR primers."  
/db\_xref="GDB:4580073"  
complement (117712, 117836)  
/standard.name="D6s2041"  
/note="SHGC-10364; The location is between each flanking  
site of PCR primers."  
/db\_xref="GDB:735115"  
join (117792, 118849, 119161, 119271, 119544, 119652,  
121626, 121682)  
/gene="TUBB"  
complement (join (117792, 118849, 119161, 119271,  
119544, 119652, 121626, 121682))  
/gene="TUBB"  
complement (join (117792, 118849, 119161, 119271,  
119544, 119652, 121626, 121682))  
/gene="TUBB"  
/product="Beta-tubulin"  
/codon.start=1  
/protein.id="BAB63321.1"  
/db\_xref="GI:15277228"  
/translation="MREIVAIQAGCCNQIGAFWEVISEDHIDPTGTHDSIDQL  
KRISVYNEATIGKIVPRALIVLEPEGMDSVSGPQIFRDNVVFGSGGNWA  
DRIETGAEALVDVLRKASCDLQFOLTHSGGSGTSGMGLDISKREYP  
DRIMFESVSPKSDTVVEPVNATLSVQLVNDETCINEMALFVDCFTLKT  
TPYGGDNLHVSAATMGVTTCTPRPQIMNLAKIVNWPFRRLPEPMGEPILTSR  
GSQQRALIVPELTQVFDKAKMMAACDPHGRILYTAAYFRMRMKEVDEGMATVQ  
NKSSYFVEMIPNNKTAACDIPRGKMAVFIQNSTALQSLFKSISEFTMFRK  
AFLEWYEGMDMEFTEASNNMDLVSEYQVQDAVAEEEDFGSEAESEA"  
117792, 118849  
/gene="TUBB"  
/number=4  
complement (117961, 118093)  
/gene="TUBB"  
/note="SGC32809; The location is between each flanking site  
of PCR primers."  
/db\_xref="GDB:4584619"  
119161, 119271  
/gene="TUBB"  
/number=3  
119544, 119552  
/gene="TUBB"  
/number=2  
121626, 121682  
/gene="TUBB"  
/number=1  
127010, 12364  
/gene="KIAA0170"  
127010, 127148  
/gene="KIAA0170"  
/number=1  
join (127013, 127148, 128005, 128385, 128471, 128540,

Query Match 29.7%; Score 89.4; DB 9; Length 200000;  
Best Local Similarity 73.7%; Pred. No. 6.4e-14;  
Matches 132; Conservative 0; Mismatches 36; Indels 11; Gaps 1;

128834, 130314, 130462, 130521, 130684, 130776,  
133831, 134625, 134734, 134804, 136089, 138566,  
138650, 138769, 138901, 139083, 139310, 139435,  
139524, 139634, 141540, 141707)  
/gene="KIAA0170"  
/note="Homologue to Drosophila photoreceptor protein  
calphactin"  
/codon.start=1  
/protein.id="BAB63322.1"  
/db\_xref="GI:15277229"  
/translation="MEPTQAIWDVDEBEETSSSRKRVNPGRIHFGAGBE  
KDFPLHGXNVGRMDPCVALPFSISKHAIETLAKAPILDDCGSLNQTQIR  
PKVYSPGVSHRLDQELIFADLQYHRLVDPFVSGLPTVBEPTVQSETPQ  
RLIAEDSEEDVFLSERRWKRSRTSSSVIPESEDEBSFVLGAPPPAFNLS  
DTVEEGQOPATEBASAAARGATVEAKGSEAVTVEIQEKQPIVKERNDTKVR  
GAGGVPAVITLERSPPEDSDPDVDSDRPRPARVLRADPFGIDTDLE  
EERIPATVIVPMKRIKIFHGVGTGRGAGLHLOBQSGSDTVEBGRAPAVPLE  
KQSAWVINDTDDEEVSAALTLAKESQPAIMWRDAEDMPQVVLQNSQITTE

RESULT 9  
AB100084  
LOCUS AB100084 300050 bp DNA linear PRI 28-JUN-2003  
DEFINITION Pan troglodytes DNA, major histocompatibility complex class I  
region, section 3/6.  
ACCESSION AB100084 BA000041  
VERSION AB100084.1 GI:32127780  
KEYWORDS  
SOURCE  
ORGANISM  
Pan troglodytes (chimpanzee)  
Pan troglodytes  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Pan.  
REFERENCE  
AUTHORS  
1 Anzai, T., Shitina, T., Kimura, N., Yanggiya, K., Kohara, S.,  
Shigenari, A., Yamagata, T., Kulski, J. K., Naruse, T. K., Fujimori, Y.,  
Fukuzumi, Y., Yamazaki, M., Tashiro, H., Iwamoto, C., Unehara, Y.,  
Imanishi, T., Meyer, A., Ikeo, K., Gojobori, T., Bahram, S. and Inoko, H.  
Comparative sequencing of human and chimpanzee MHC class I regions  
unveils insertions/deletions as the major path to genomic  
divergence  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 100 (13), 7708-7713 (2003)  
MEDLINE 22709134  
PUBMED 12799463  
REFERENCE  
AUTHORS  
2 (bases 1 to 300050)  
Anzai, T. and Shitina, T.  
TITLE Direct Submission  
SUBMITTED (15-JAN-2003) Tatsuya Anzai, Tokai University School of  
Medicine, Department of Molecular Life Science 2; Bohseidai,  
Isehara, Kanagawa 259-1193, Japan  
E-mail: anzai@is.icc.u-tokai.ac.jp, Tel: 81-463-93-1121 (ex. 2652),  
Fax: 81-463-94-8884  
FEATURES  
source  
1..300050  
Location/Qualifiers  
/organism="Pan troglodytes"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9598"  
1..6411  
misc\_feature



/note="Library: RPCI-43, Clone: 148D08"  
complement (1833, .7664)  
/gene="TFIIH"  
complement (1833, .2134)  
/gene="TFIIH"  
/number=14  
complement (join(1962, .2134,2558, .2636,2788, .2835,  
3488, .3618,3800, .3932,4179, .4262,4447, .4515,4677, 4788,  
5005, .5093,5185, .5281,5883, .6014,6407, .6511,6773, .6909))  
/gene="TFIIH"  
/codon\_start=1  
/product="transcription factor II H"  
/protein\_id="BAC78171.1"  
/db\_xref="GI:32127781"  
/translation="WESTERGLNRLVHLQCRNTQEPFGISPGVLDRLYGHPATCIAV  
FRELPLAKWMMMLFLEQPLPQAAVALWKEFSKAOESTGLSGRLMWTOLP  
GGLQGLINPFRQNLRIALLGGKASDSTQSGPDKHARDVPSLDKXAESEVVL  
HFMGSPSAVSODIQLQSAGLMKSTEGEPCISAGFOPLDTPQALMYFMLO  
YLORASRGMDIVYIISFLFQLSRSTGKQSVSGMSDILNFLQHLREGQVFSCKR  
KSRIVYPTRLAINLSGVSQAQGTTHOPGIVVETMYRLYATSEIQLALLISEM  
LYREFNNVAQVETRESVQQAISGITHQQLIHFLRFRAPVMLKQTPVLPPTTDIR  
VKELEERDLRFTFEGVLNPFQVDFELLHAHRELGLVFENSARKLMVVTAGSD  
VKRFPMKXKSS"  
complement (2558, .2636)  
/gene="TFIIH"  
/number=13  
complement (2788, .2835)  
/gene="TFIIH"  
/number=12  
complement (3488, .3618)  
/gene="TFIIH"  
/number=11  
complement (3800, .3932)  
/gene="TFIIH"  
/number=10  
complement (4179, .4262)  
/gene="TFIIH"  
/number=9  
complement (4447, .4515)  
/gene="TFIIH"  
/number=8  
complement (4677, .4788)  
/gene="TFIIH"  
/number=7  
complement (5005, .5093)  
/gene="TFIIH"  
/number=6  
complement (5185, .5281)  
/gene="TFIIH"  
/number=5  
complement (5883, .6014)  
/gene="TFIIH"  
/number=4  
complement (6407, .6511)  
/gene="TFIIH"  
/number=3  
6412, .183556  
/note="Library: RPCI-43, Clone: 158N24"  
complement (6773, .6912)  
/gene="TFIIH"  
/number=2  
complement (7543, .7664)  
/gene="TFIIH"  
/number=1  
complement (16929, .27490)  
/gene="DDR"  
complement (join(16929, .17069,17188, .17337,17918, .18152,  
18638, .18838,19076, .19203,19352, .19596,20702, .20812,  
21566, .21711,22790, .22941,23050, .23145,23671, .23917,  
24026, .24212,24743, .24842,25102, .25249,26790, .27006,  
27210, .27312,27406, .27490))  
/gene="DDR"  
/codon\_start=1

/product="receptor tyrosine kinase"  
/protein\_id="BAC78172.1"  
/db\_xref="GI:32127782"  
/translation="WGPALSSLLILLIVASGDADMKGHPPPAKCRVALGKQDRTIPD  
SDISASSMSDSTLARSSDDGAMCPGASVPPEEEYLOVDLQRLHLVALVYQGRH  
AGGLGKERSRSYRLRYSDRGMWGMKRMQGEVYISGNEDEGVLDLGPVYARLY  
RFPYADRVMSVCLRVLEYGLMRDGLSYAPVGQNTYLSAAVLYNDSTYDGTVGG  
LOYGGLQGLADGVGLDFFRSQSLRWVPGVDVYGMSSHSFGVYEMEPFDLRF  
QAMQYHCNNHTTLGARLPGVYCFRRRPPAMAMGEPRRNLGSLDPPRARAYPL  
GGRVARFLQCRFLFAGFWLFSLSFSDVNNSSPALGGTFPPAPWMPGPPPTNS  
SLELEPRQCPVAAEAGSPPTAILIGCVAILILLILLITALLMLRLHMRLLSKERY  
LESELTVLSPVGDITILINRRGPREPYPGPRGPPSPACVPPGSLILSNPA  
YRLIATYAPRPGGPPTPAMAKPTNQAAYSGMYMEPEKGAFLPPPPNSVPHA  
EADIYTLQGTGNTYANPALPPGAVDGPFRVYPPRRRLAFKRLGEGCGRYHLE  
VDSFDVLVDLPFLVNRKGFLLVAVKLRFDAIKANNDLKVKIMSLKDNITIR  
LLGVQDDPPLCMITDVNENDLQVFSANGLERKAAGAGDQAAGPTISYPMIL  
HYAQIASGCVYALINLFVHDLATRNCLVGNFTIKIADFGSMRLYAGDYVQGR  
AVLPFRMWGCIMLGKFTTASDYWAFVTLMEYLMICRAQPGQLTDEQVIEAGAF  
FRDQGRQYVLSRPACPGCLYELMLRCMSRESBGRPPRSQHLRLADALNTV"  
complement (<16929, .17069)  
/gene="DDR"  
/number=17  
complement (17188, .17337)  
/gene="DDR"  
/number=16  
complement (17918, .18152)  
/gene="DDR"  
/number=15  
complement (18638, .18838)  
/gene="DDR"  
/number=14  
complement (19076, .19203)  
/gene="DDR"  
/number=13  
complement (19352, .19596)  
/gene="DDR"  
/number=12  
complement (20702, .20812)  
/gene="DDR"  
/number=11  
complement (21546, .21711)  
/gene="DDR"  
/number=10  
complement (22790, .22941)  
/gene="DDR"  
/number=9  
complement (23050, .23145)  
/gene="DDR"  
/number=8  
complement (23671, .23917)  
/gene="DDR"  
/number=7  
complement (24026, .24212)  
/gene="DDR"  
/number=6  
complement (24743, .24842)  
/gene="DDR"  
/number=5  
complement (25102, .25249)  
/gene="DDR"  
/number=4  
complement (26790, .27006)  
/gene="DDR"  
/number=3  
complement (27210, .27312)  
/gene="DDR"  
/number=2  
complement (27406, .27490)  
/gene="DDR"  
/number=1  
171051, .172310  
/gene="EX-1"  
join(171051, .17260,171373, .171633)



/gene="IEX-1"  
 /note="PRG1  
 expressed in pancreatic carcinoma"  
 /codon\_start=1  
 /protein\_id="BAC78173.1"  
 /db\_xref="GI:32127783"  
 /translation="MCHSRSCHPMTLILQAPAPSPNPGRRSGPEIFTPDLPPEP  
 AAPAGRPASRGHRRSRVYPCVVRRLPVEEPNPAKRLIFLLTIFPCOIMAE  
 EGYPAPEPDEDAENASLAPTPVSPVLEPNNLSESDVALDSTLQHPAAF"  
 <171051..171260  
 /gene="IEX-1"  
 /note="PRG1"  
 /number=1  
 171373..172310  
 /gene="IEX-1"  
 /note="PRG1"  
 /number=2  
 172876..187775  
 /gene="HLOT1"  
 gene

Query Match 29.2%; Score 87.8; DB 9; Length 300050;  
 Best Local Similarity 73.2%; Pred. No. 1,7e-13;  
 Matches 131; Conservative 0; Mismatches 37; Indels 11; Gaps 1;

QY 87 TGTGATATTGATATATATATAGTTTGTTCATAGTTCCTGCTTATTAACCCCTC 146  
 DB 110384 TGTGCTATTATATATATATATAGTTTGTTCATAGTTCCTGCTTATTAATCCATT 110443  
 QY 147 ACCCTGTTACAGTCATTTGTTATAGGTGATGTTAGGCTCAGAACAAACTC- 205  
 DB 110444 GCCCTGTTACCGCTGTGTTGTAAGTTGGTGCTTAAGCTCAGAACAGGTCTCA 110503  
 QY 206 -----TCTCTCTACCTTCTCCAGCCCTCTCTCTCTGCGACCTCATTTCTCC 254  
 DB 110504 ATTAACGCAATCTCTCATTTCTCTCTGCTCATTTCACTGCTCTTATCTCTCC 110562

RESULT 10  
 HSP20F2/c 481 bp DNA linear PRI 23-AUG-1996  
 LOCUS H sapiens flow-sorted chromosome 6 HindIII fragment, SC6PA20F2.  
 DEFINITION Z79016  
 VERSION Z79016.1 GI:1508294  
 KEYWORDS Anonymous marker; single read.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 Mungall, A.J., Huckle, E., Langford, C., Ross, M.T. and Rice, C.M.  
 TITLE Direct Submission  
 JOURNAL Submitted (22-AUG-1996) The Sanger Centre, Wellcome Trust Genome  
 Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail contact:  
 humquery@sanger.ac.uk  
 Vector: pBS1ISK+.

COMMENT  
 FEATURES  
 source  
 1..481  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="6"  
 /clone="SC6PA20F2"  
 /sex="female"  
 /tissue\_type="EBV lymphoblastoid cell line"  
 /clone\_lib="SC6PA"  
 /dev\_stage="adult"  
 /note="The estimated purity of the flow-sorted chromosome  
 6 library is >97%"

ORIGIN  
 Query Match 27.8%; Score 83.8; DB 9; Length 481;  
 Best Local Similarity 76.3%; Pred. No. 2.9e-12;  
 Matches 103; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 87 TGTGATATTGATATATATATAGTTTGTTCATAGTTCCTGCTTATTAACCCCTC 146  
 DB 234 TGTGCTATTATATATATATATAGTTTGTTCATAGTTCCTGCTTATTAATCCATT 175  
 QY 147 ACCCTGTTACAGTCATTTGTTATAGGTGATGTTAGGCTCAGAACAAACTC 206  
 DB 174 GCCCTGTTACCGCTGTGTTGTAAGTTGGTGCTTAAGCTCAGAACAGGTCTCT 115  
 QY 207 CTCTCTACCTTCTC 221  
 DB 114 GACCTTCTCTGCCC 100

RESULT 11  
 HSDJ93K22 113804 bp DNA linear PRI 04-MAR-2003  
 LOCUS Human DNA sequence from clone RPI-93K22 on chromosome 6q14.1-15.3,  
 complete sequence.  
 DEFINITION AL050333  
 ACCESSION AL050333.18 GI:6911641  
 VERSION HTG.  
 KEYWORDS Homo sapiens (human)  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 113804)  
 Tracey, A.  
 Direct Submission  
 Submitted (04-MAR-2003) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humquery@sanger.ac.uk  
 On Feb 7, 2000 this sequence version replaced gi:6807622.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 The following abbreviations are used to associate primary accession  
 numbers given in the feature table with their source databases:  
 Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information  
 on the WORMPEP database can be found at  
 http://www.sanger.ac.uk/projects/C\_elegans/wormpep  
 This sequence was generated from part of bacterial clone contigs of human  
 chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping  
 Group. Further information can be found at  
 http://www.sanger.ac.uk/HGP/Chr6  
 ----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: humquery@sanger.ac.uk  
 -----

This sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >=  
 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest, except on the rare  
 occasion of the clone being a YAC.  
 RPI-93K22 is from the library RPI-1 constructed by the group of  
 Piter de Jong. For further details see  
 http://www.chori.org/bacpac/home.htm  
 VECTOR: pCYPAC2.

COMMENT  
 FEATURES  
 source  
 1..113804  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="6"  
 /map="q14.1-15.3"



ORIGIN /clone="RP1-93K22"  
/clone\_lib="RPCT-1"

Query Match 27.8%; Score 83.8; DB 9; Length 113804;  
Best Local Similarity 76.3%; Pred. No. 2,2e-12;  
Matches 103; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 87 TGTGATATATATATATATATATAGCTTTTGTCCATAGTTCCTGCTTAATAAACCCCTC 146  
DB 112699 TGTGATATATATATATATATATAGCTTTTGTCCATAGTTCCTGCTTAATAAACCCCTC 112758

QY 147 ACCCTTGTTACAGCATTGTTGTTAAGTGGATGCTTGGCTCAGAACCAAACTCT 206  
DB 112759 GCCCTTGTTACAGCATTGTTGTTAAGTGGATGCTTGGCTCAGAACCAAACTCT 112818

QY 207 CTCCTTCACCTTCTC 221

DB 112819 GACCTTCCTCCTGCC 112833

RESULT 12  
AC025610/c 170877 bp DNA linear HTG 30-MAR-2000  
LOCUS Homo sapiens chromosome 6 clone RP11-413M1 map 6, WORKING DRAFT  
DEFINITION  
SEQUENCE, 36 unordered pieces.  
AC025610  
AC025610.2 GI:7342184  
VERSION HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
KEYWORDS Homo sapiens (human)  
SOURCE  
ORGANISM

REFERENCE  
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bada, F., Boguslavsky, L., Bouckhagalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A., Choquet, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J.S., Dodge, S., Domingo, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Labocque, K., Lamazares, R., Landers, T., Lehotzky, V., Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Margulis, N., McCarthy, M., McEwan, P., McGuirk, A., McKernan, K., McPherson, J., Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mlenka, V., Morrow, J., Murphy, T., Naylor, T., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Olivari, T.M., Oliver, J., Peterson, K., Pierre, N., Plesniak, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Teasdale, S., Theodore, J., Tirtrell, A., Travers, M., Triggiano, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (12-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 30, 2000 this sequence version replaced gi:7229942.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/JM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: MIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence.submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L8333

Center clone name: 413.M.1  
----- Summary Statistics  
Sequencing vector: M13; M77815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 147503 bases at least Q40  
Consensus quality: 159832 bases at least Q30  
Consensus quality: 164267 bases at least Q20  
Insert size: 165000; agarose-fp  
Insert size: 167377; sum-of-ctnigs  
Quality coverage: 2.8 in Q20 bases; agarose-fp  
Quality coverage: 3.1 in Q20 bases; sum-of-ctnigs  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 36 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1	1682: contig of 1682 bp in length
1683	1782: gap of 100 bp
1783	2939: contig of 1157 bp in length
2940	3039: gap of 100 bp
3040	4951: contig of 1512 bp in length
4952	5051: gap of 100 bp
5052	6630: contig of 1579 bp in length
6631	6730: gap of 100 bp
6731	8341: contig of 1611 bp in length
8342	8441: gap of 100 bp
8442	10096: contig of 1655 bp in length
10097	10196: gap of 100 bp
10197	11547: contig of 1351 bp in length
11548	11647: gap of 100 bp
11648	13051: contig of 1404 bp in length
13052	13151: gap of 100 bp
13152	14432: contig of 1281 bp in length
14433	14532: gap of 100 bp
14533	16465: contig of 1933 bp in length
16466	16565: gap of 100 bp
16566	18763: contig of 2198 bp in length
18764	18863: gap of 100 bp
18864	20876: contig of 2013 bp in length
20877	20976: gap of 100 bp
20977	23264: contig of 2288 bp in length
23265	23364: gap of 100 bp
23365	25715: contig of 2351 bp in length
25716	25815: gap of 100 bp
25816	29287: contig of 3472 bp in length
29288	29387: gap of 100 bp
29388	32438: contig of 3051 bp in length
32439	32538: gap of 100 bp
32539	33835: contig of 3257 bp in length
33836	35935: gap of 100 bp
35936	39035: contig of 3100 bp in length
39036	39135: gap of 100 bp
39136	40320: contig of 1185 bp in length
40321	40420: gap of 100 bp
40421	45654: contig of 5234 bp in length
45655	45754: gap of 100 bp
45755	50391: contig of 4637 bp in length
50392	50491: gap of 100 bp
50492	54449: contig of 3958 bp in length
54449	54549: gap of 100 bp
54550	59425: contig of 4876 bp in length
59426	59525: gap of 100 bp
59526	65171: contig of 5646 bp in length
65172	65271: gap of 100 bp
65272	68021: contig of 3750 bp in length
68022	69121: gap of 100 bp
69122	75111: contig of 5990 bp in length
75112	75211: gap of 100 bp



```
misc_feature 1..1682 /note="assembly_fragment"
misc_feature 1783..2939 /note="assembly_fragment"
misc_feature 3040..4951 /note="assembly_fragment"
misc_feature 5052..6630 /note="assembly_fragment"
misc_feature 6731..8341 /note="assembly_fragment"
misc_feature 8442..10096 /note="assembly_fragment"
misc_feature 10197..11547 /note="assembly_fragment"
misc_feature 11648..13051 /note="assembly_fragment"
misc_feature 13152..14432 /note="assembly_fragment"
misc_feature 14533..16465 /note="assembly_fragment"
misc_feature 16566..18763 /note="assembly_fragment"
misc_feature 18864..20876 /note="assembly_fragment"
misc_feature 20977..23264 /note="assembly_fragment"
misc_feature 23365..25715 /note="assembly_fragment"
misc_feature 25816..29287 /note="assembly_fragment"
misc_feature 29388..32438 /note="assembly_fragment"
misc_feature 32539..35835 /note="assembly_fragment"
misc_feature 35936..39035 /note="assembly_fragment"
misc_feature 39136..40320 /note="assembly_fragment"
misc_feature clone_end:Spe vector_side:right"
misc_feature 40421..45654 /note="assembly_fragment"
misc_feature 45755..50391 /note="assembly_fragment"
misc_feature 50492..54449
```

RESULT	13
LOCUS	AC021655
DEFINITION	Homo sapiens chromosome 3 clone RP11-153X19, WORKING DRAFT SEQUENCE, 26 unordered pieces.
ACCESSION	AC021655
VERSION	AC021655.21 GI:21431092
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 143133) Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,

AC021655 143133 bp DNA linear HTG 21-SEP-2002  
 Homo sapiens chromosome 3 clone RP11-153K19, WORKING DRAFT  
 SEQUENCE, 26 unordered pieces.  
 AC021655  
 AC021655.21 GI:21431092  
 HTG: HTGS PHASE1, HTGS\_DRAFT.  
 Homo sapiens (human)  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 143133)  
 Muzny, D.M., Adams, C., Adio-oduola, B., Ali-ozman, F.R., Allen, C.,  
 Alsbrooks, S.L., Anaratunga, H.C., Are, J.R., Ayale, M., Banks, T.,  
 Barbieri, J., Benton, U., Bimoge, K., Blankenburg, K., Bonnin, D.,  
 Bouck, J., Bowse, S., Briviera, M., Brown, S., Brown, K., Bryan, N.P.,  
 Buahy, C., Burch, P., Burkett, C., Butrell, K.L., Byrd, N.,  
 Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,  
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,  
 Cleveland, C.D., Cox, C., Coyle, M.D., Dahorne, S.R., David, R.,  
 Davila, M.L., Davis, C., Davy-carroll, L., Dedereich, D.A.,  
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,  
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,  
 Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escoto, M.,  
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,  
 Gabel, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,  
 Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,  
 Harris, C., Harris, K., Hart, M., Havlik, P., Haws, A., Hernandez, J.,  
 Hernandez, O., Hodgson, A., Hughes, M., Holloway, C., Hollins, B.,  
 Homs, F., Howard, S., Huber, J., Huliy, S., Hume, J., Jackson, L.E.,  
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,  
 Jacobson, E., Kelly, S., Khan, U., King, L., Korwah, J., Kovar, C.,  
 Kratovic, J., Kureishi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,  
 Li, U., Li, Z., Lichtharge, O., Lieu, C., Liu, T., Liu, W., Louisaged, H.,  
 Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
 Maheshwari, M., Mapa, P., Martin, R., Martinale, A., Martinez, E.,  
 Massey, E., Maunine, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M.,  
 Miner, G., Miner, Z., Mitchell, T., Mohabhat, K., Morgan, M., Morris, S.,  
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
 Nguyen, N., Nicolson, E., Nwokweto, S., Ogun, M., Okunolu, G.,  
 Oranumbe, N., Orfield, R., Pace, A., Payton, J., Peery, J., Perez, L.,  
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,  
 Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G.,  
 Scheraga, S., Scott, G., Shen, H., Shoochattai, N., Sison, I.,  
 Sodergren, E., Sonakhe, T., Sparks, A., Stanley, H., Stone, H.,  
 Sutton, A., Swalek, A., Tabor, P., Tameris, A., Tamejisa, K., Tang, H.,  
 Tansley, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
 Umami, K., Vazquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,  
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
 Williams, G., Williamson, A., Wleczyk, R., Woodson, S., Wolley, K.,



TITLE	JOURNAL	REFERENCE	AUTHORS	TITLE	JOURNAL
Unpublished				Wu, C., Mu, Y., Mu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstock, G. and Gibbs, R.	
2 (baees 1 to 143133)					
Worley, K.C.					
Direct Submission					
Submitted (19-JAN-2000)					
of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA					
3 (baees 1 to 143133)					
Worley, K.C.					
Direct Submission					
Submitted (21-SEP-2002)					
of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA					
On Jun 17, 2002 this sequence version replaced gi:20335448.					



TITLE	JOURNAL	REFERENCE
AUTHORS	JOURNAL	
COMMENT		
		Dugan,A.L., Ding,Y., Dink,H.H., Douthwaite,C.J., Draper,H., Elhaj,C., Escoto,C., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Foster,P., Frantz,P., Falls,T., Ferraguto,D., Flagg,N., Ford,U., Garza,N., Gill,R., Gorrell,U.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homel,P., Howard,S., Huber,J., Hulky,S., Hune,U., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Jordah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtagse,O., Lieu,C., Lien,J., Liu,W., Lundsegged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Marindale,A., Martinez,E., Massey,E., Mashiney,E., McLeod,M.S., Meador,M., Mei,G., Metker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkenkwo,S., Ogih,M., Okunnu,G., Oragunye,N., Oviedo,R., Pace,A., Peyton,B., Peery,J., Peters,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojoudokan,I., Rolfe,M., Rutiz,S., Saverly,G., Scherer,S., Scott,G., Shen,H., Shochetari,N., Sisson,I., Sodergren,E., Sonalike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svarek,A., Tabors,P., Tamerias,A., Tameria,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmari,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wlezyk,R., Wodden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G. and Gibbs,R.
	Unpublished	
	2 (bases 1 to 163862)	
	Direct Submission	
	Submitted (09-AUG-2001)	Human Genome Sequencing Center, Department
	of Molecular and Human Genetics, Baylor College of Medicine, One	
	Baylor Plaza, Houston, TX 77030, USA	
	On Nov 1, 2001 this sequence version replaced gi:16445074.	

FEATURES  
source

```

*
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html)
*
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contig. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 163862: contig of 163862 bp in length.
* Location/Qualifiers
* 1..163862
* /organism="Homo sapiens"
* /mol_type="genomic DNA"

```

```

/db xref="taxon:9606"
/chromosome="3q"
/clone="RP11-776D7"
ORIGIN

```

Query Match	27.4%	Score 82.6;	DB 2;	Length 163862;	
Best Local Similarity	75.2%	Pred. No. 4.6e-12;			
Matches 103;	Conservative	0;	Mismatches 34;	Indels 0;	Gaps 0

[illegible]

```

OY      142 CCCGACCCCTGTACAGTCATTGTTATTAAGTTGATGGTTTAAGGCTCAGAACCA 201
      || | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      4761 CCATAGCCCTGTTCGATCATTTGTATTAATGTTGATATGTTAAGCATCAGNAACA 4822

```

QY	202	ACCTCTCTCTACCTT	218
Db	4821	AACTCTCTGACCTCT	4837

LOCUS	AC026559	165082 bp	DNA	linear	HTG 24-AUG-2002
DEFINITION	Homo sapiens chromosome 4 clone RP11-776D7 map 4, WORKING DRAFT				
SEQUENCE	SEQUENCE, 8 unordered pieces.				
ACCESSION	U000077				

VERSION	AC026559.2	GI:7467245
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT	
SOURCE	Homo sapiens (human)	

```

Euteleostomi
Chordata; Vertebrata; Euteleostomi
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 /homo 1 ka 100000

```

AUTHORS	Blirren, B., Linton, L., Nusbaum, C. and Lander, E
TITLE	Homo sapiens chromosome 4, clone RP11-776D7
FORMAT	FASTA

## AUTHORS

AUTHORS	TITLE
Barren, B., Linton, L., Nusbach, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F., Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G., Campiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L., Howland, J.C., Illey, I., Johnson, R., Jones, C., Kam, L., Karakas, A., Klein, J., LeRoque, K., Lamazars, R., Landers, T., Lehoczy, U., Levine, R., Lien, C., Liu, G., Locke, K., MacDonald, P., Margie, N., McCarthy, M., McEwan, P., McGuirk, A., McKernan, K., McPheters, R., Meldrum, T., Meneus, L., Mihova, C., Miranda, C., Mlenga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, T.M., Oliver, J., Peterson, C., Pierre, N., Piatani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Sudraman, A., Talamas, J., Teifate, S., Theodore, J., Tirrell, A., Travers, M., Trifilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.	Submitted (22-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA 3 (Bases 1 to 165082)
REFERENCE	AUTHORS
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F., Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G., Campiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L., Howland, J.C., Illey, I., Johnson, R., Jones, C., Kam, L., Karakas, A., Klein, J., LeRoque, K., Lamazars, R., Landers, T., Lehoczy, U., Levine, R., Lien, C., Liu, G., Locke, K., MacDonald, P., Margie, N., McCarthy, M., McEwan, P., McGuirk, A., McKernan, K., McPheters, R., Meldrum, T., Meneus, L., Mihova, C., Miranda, C., Mlenga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, T.M., Oliver, J., Peterson, C., Pierre, N., Piatani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Sudraman, A., Talamas, J., Teifate, S., Theodore, J., Tirrell, A., Travers, M., Trifilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.	Submitted (22-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA 3 (Bases 1 to 165082)



Klein, J., Lapocque, K., Lamazez, R., Landers, T., Leboczky, J.,  
 Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marcus, N.,  
 McCarthy, M., McKean, P., McGuirk, A., McKernan, K., McNeelers, R.,  
 Meldrum, J., Menuez, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,  
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,  
 O'Neill, D., Oliver, T.M., Oliver, J., Petersen, K., Pierre, R.,  
 Pisani, C., Pollars, A., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
 Stange-Bromann, N., Stojanovic, N., Subramanian, A., Tamas, J.,  
 Testeys, S., Theodore, J., Tirrelli, A., Travers, M., Triggilio, J.,  
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.U.,  
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.  
 Direct Submissions  
 Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Apr 6, 2000 this sequence version replaced gi:17283272.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Comugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: March 12, 2004, 12:26:39 / Search time 120.53 seconds

(without alignments)  
10609.014 Million cell updates/sec

Title: US-09-966-880A-10\_COPY\_6264\_6564

Perfect score: 301 1 gctcaagaagcacttagta.....cttcacaaagcgctcaag 301

Sequence: 1 gctcaagaagcacttagta.....cttcacaaagcgctcaag 301

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 200000000Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_29Jan04:\*  
1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002as:\*  
7: geneseqn2003as:\*  
8: geneseqn2003bs:\*  
9: geneseqn2003cs:\*  
10: geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	301	100.0	6564	3 AAC55314	AAC55314 Human act
2	73.4	24.4	355	4 AAI89801	AAI89801 Human pol
3	71.8	23.9	612	5 ABV58443	ABV58443 Human pro
4	70.4	23.4	740	4 AAI96626	AAI96626 Human neu
5	68.8	22.9	864	4 AAI86523	AAI86523 Human pol
6	68.8	22.9	2102	4 AAH15837	AAH15837 Human CDN
7	67	22.3	875	4 AAH82024	AAH82024 Human imm
8	67	22.3	35465	4 AAF54723	AAF54723 Nucleotid
9	67	22.3	35465	7 ABT17380	ABT17380 Human IG
10	67	22.3	36991	7 ABT17382	ABT17382 Human IG
11	66.8	22.2	621	6 ABN62372	ABN62372 Human can
12	65.8	21.9	452	4 AAK62248	AAK62248 Human imm
13	65.8	21.9	1524	4 AAL02965	AAL02965 Human rep
14	65.8	21.9	1524	4 ABA07734	ABA07734 Human ova
15	65	21.6	452	8 ACH26427	ACH26427 Human act
16	65	21.6	2141	4 AAS02525	AAS02525 Human sec
17	65	21.6	2510	6 AAD31155	AAD31155 Human Alz
18	64.4	21.4	13069	4 AAK36350	AAK36350 Human car
19	64.4	21.4	13069	4 AAK74980	AAK74980 Human imm
20	64.4	21.4	13069	4 ADA77044	ADA77044 Human car
21	64.4	21.4	32203	4 AAS36349	AAS36349 Human car
22	64.4	21.4	32203	9 ADE47043	ADE47043 Human car
23	64.4	21.4	47513	4 AAK74979	AAK74979 Human imm

24	63.4	21.1	21313	4 AAK82710	AAK82710 Human imm
25	63.4	21.1	222930	6 ABK84349	ABK84349 Human CDN
26	62.6	20.8	110000	6 ABX08336_13	Continuation (14 o
27	62.6	20.8	110000	7 ACE42745_0	ACE42745 Human ALM
28	62.4	20.7	1480	7 ADC85848	ADC85848 Human GPC
29	62	20.6	957	3 AAC74307	AAC74307 Human sec
30	62	20.6	1198	4 AAI92788	AAI92788 Human pol
31	62	20.6	1733	3 AAC76894	AAC76894 Human ORF
32	61.8	20.5	736	4 AAK69223	AAK69223 Human imm
33	61.4	20.4	48727	4 AAK67375	AAK67375 Human imm
34	60.4	20.1	148567	7 ABSS5500	ABSS5500 Gene enco
35	60.4	20.1	148567	8 ACA62841	ACA62841 Human kin
36	59.2	19.7	967	6 ABZ11514	ABZ11514 Human pol
37	58.2	19.3	39776	7 AAL51353	AAL51353 Human sec
38	57.4	19.1	416	8 ACH48924	ACH48924 Human leu
39	57.4	19.1	585	4 AAI10627	AAI10627 Human CDN
40	57	18.9	2626	4 AAH17370	AAH17370 Human CDN
41	56.6	18.8	5879	6 ABL32268	ABL32268 Human imm
42	55.8	18.5	174493	7 ACA61395	ACA61395 Novel hum
43	55.8	18.5	174493	9 AAD59937	AAD59937 Human kin
44	55.4	18.4	381	6 ABV95367	ABV95367 Human pan
45	55.4	18.4	546	3 AAH31133	AAH31133 Human col

## ALIGNMENTS

RESULT 1	
AAC55314	
ID AAC55314 standard; DNA; 6564 BP.	
XX AAC55314;	
XX	
DT 05-FEB-2001 (first entry)	
XX	
DE Human activation-induced cytidine deaminase genomic DNA SEQ ID NO:10.	
XX	
OS Homo sapiens.	
XX	
PN WO200058480-A1.	
XX	
PD 05-OCT-2000.	
XX	
PF 28-MAR-2000; 2000WO-JP001918.	
XX	
PR 29-MAR-1999; 99JP-00087192.	
XX	
PR 24-JUN-1999; 99JP-00178999.	
XX	
PR 27-DEC-1999; 99JP-00371382.	
XX	
PA (NIBS) JAPAN TOBACCO INC.	
XX	
PI (HONJ/) HONJO T.	
XX	
PI Honjo T, Muramatsu M;	
XX	
DR WPI; 2000-611715/58.	
XX	
PT Nucleic acid encoding activation induced cytidine deaminase, useful as a	
XX	
PT target for drug development for immune-related diseases including	
XX	
PT allergies.	
XX	
PS Claim 17, Page 145-150; 174pp; Japanese.	
XX	







PR 13-DEC-2000; 2000US-0255281P.  
XX  
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX  
XX Schlegel R, Endege WO, Monahan JB;  
XX  
XX WPI; 2001-662795/76.  
XX  
PT Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PT for detecting presence of prostate cancer, stage of prostate cancer.  
XX  
PS Claim 1; Page 11218; 11750pp; English.  
XX  
XX The invention relates to an isolated nucleic acid molecule (i) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
CC specification or its complement. (i) is useful for: (a) assessing whether  
CC a patient is afflicted with prostate cancer; (b) monitoring the  
CC progression of prostate cancer in a patient; (c) assessing the efficacy  
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing  
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)  
CC determining whether prostate cancer has metastasized in a patient; (h)  
CC assessing the aggressiveness or indolence of prostate cancer in a patient  
CC ; (i) is also useful as a pharmacodynamic or pharmacogenomic marker  
XX  
SQ Sequence 612 BP; 157 A; 125 C; 124 G; 206 T; 0 U; 0 Other;  
XX  
Query Match 23.9%; Score 71.8; DB 5; Length 612;  
Best Local Similarity 69.8%; Pred. No. 1.3e-11;  
Matches 97; Conservative 0; Mismatches 42; Indels 0; Gaps 0;  
XX  
QY 98 GATATATATATAGATTTTGTCCATAGTTCCTGCTTATATAACCCCTCACCTTGTAC 157  
XX  
DB 204 GATACAGATATAGATTTTGTCCAGTCTTGGCTCACAACCTCATAGCCCTTGTAT 145  
XX  
QY 158 AGTCATTTTATATAGATGATGTTTAGGCTCAGAAACAACCTCTCTTCACCT 217  
XX  
DB 144 AGTTTGTGTATATAGTGGTGTGTGCGGCTCAGAAATGATCTCTGACCTTC 85  
XX  
QY 218 TCTCCAGCCCTCCTGCTC 235  
XX  
DB 84 TGTCCTTCTTCCACCTGTC 66  
XX  
RESULT 4  
AA196626/c  
ID AA196626 standard; cDNA; 740 BP.  
XX  
XX AA196626;  
AC  
XX  
DT 13-NOV-2001 (first entry)  
XX  
DE Human neuroblastoma expressed polynucleotide SEQ ID NO 2701.  
XX  
XX Human neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.  
XX  
OS Homo sapiens.  
XX  
XX WO200166719-A1.  
XX  
XX 13-SEP-2001.  
XX  
XX 02-MAR-2001; 2001WO-JP001629.  
XX  
XX 07-MAR-2000; 2000JP-00159195.  
XX  
XX (CHIB-) CHIBA PREFECTURE.  
XX  
XX (HISM) HISAMITSU PHARM CO LTD.  
XX  
XX Nakagawara A;  
XX

DR WPI; 2001-565584/63.  
XX  
XX Nucleic acids originating in gene expressed in human neuroblastoma,  
PT useful as probe or primer in diagnosing prognosis of human neuroblastoma,  
PT malignancy and susceptibility indicator or tumor marker for anti-cancer  
PT agents.  
XX  
XX Claim 1; Page 1981; 2979pp; Japanese.  
XX  
XX The invention relates to novel genes (AA193926-AA197963) expressed in  
CC human neuroblastoma. The nucleic acids are applicable as a probe or  
CC primer in diagnosing the prognosis of human neuroblastoma, malignancy and  
CC susceptibility indicators or tumour markers for anti-cancer agents. The  
CC gene information for diagnosing prognosis is related to factors similar  
CC to that for N-myc and TrkA genes  
XX  
SQ Sequence 740 BP; 184 A; 122 C; 199 G; 218 T; 0 U; 17 Other;  
XX  
Query Match 23.4%; Score 70.4; DB 4; Length 740;  
Best Local Similarity 83.6%; Pred. No. 3.7e-11;  
Matches 92; Conservative 0; Mismatches 16; Indels 2; Gaps 1;  
XX  
QY 188 GCTTCAGAGAGAAAATCTCTCTCTCACTTCTTCAGCCCTCCTGTTCTTGACCTCAT 247  
XX  
DB 562 GCCCAGAGAGAAAAGTTCTCTCTGACCTTCTCTCCGCCCTTGTCTTGCGCTTTCAT 503  
XX  
QY 248 TCTTCCCTGAGGC--CACTAGAAACTGAGATCTCTTTCACAGCGG 295  
XX  
DB 502 TCTCCCCAGAGGCTACCCCATAGAAACTGAAATCCTCTTCCCAAGGCG 453  
XX  
RESULT 5  
AA186523/c  
ID AA186523 standard; cDNA; 864 BP.  
XX  
XX AA186523;  
AC  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human polynucleotide SEQ ID NO 6583.  
XX  
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
XX tissue growth factor; immunomodulatory; cancer; leukaemia;  
XX nervous system disorders; arthritis; inflammation; ss.  
XX  
OS Homo sapiens.  
XX  
XX WO200164935-A2.  
XX  
XX 07-SEP-2001.  
XX  
XX 26-FEB-2001; 2001WO-US004927.  
XX  
XX 28-FEB-2000; 2000US-00515126.  
XX  
XX 18-MAY-2000; 2000US-00577409.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Dimaenac RT;  
XX  
XX WPI; 2001-514838/56.  
XX  
XX P-PSDB; AAO06592.  
XX  
XX Isolated nucleic acids and polypeptides, useful for preventing diagnosing  
PT and treating e.g. leukemia, inflammation and immune disorders.  
XX  
XX Claim 1; SEQ ID NO 6583; 1399pp + Sequence Listing; English.  
XX  
XX The invention relates to human polynucleotides (AA179941-AA193841) and  
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The







PR	14-AUG-2000	2000US-02256566
PR	14-AUG-2000	2000US-02256678
PR	14-AUG-2000	2000US-02256687
PR	14-AUG-2000	2000US-02252700
PR	14-AUG-2000	2000US-02255477
PR	14-AUG-2000	2000US-02257578
PR	14-AUG-2000	2000US-02257588
PR	14-AUG-2000	2000US-02257598
PR	18-AUG-2000	2000US-02262719
PR	18-AUG-2000	2000US-02266819
PR	22-AUG-2000	2000US-02266868
PR	22-AUG-2000	2000US-02267182
PR	23-AUG-2000	2000US-02270092
PR	30-AUG-2000	2000US-02289548
PR	01-SEP-2000	2000US-02292878
PR	01-SEP-2000	2000US-02293438
PR	01-SEP-2000	2000US-02293448
PR	01-SEP-2000	2000US-02293458
PR	05-SEP-2000	2000US-02295058
PR	05-SEP-2000	2000US-02295138
PR	06-SEP-2000	2000US-02304378
PR	06-SEP-2000	2000US-02304388
PR	08-SEP-2000	2000US-02320808
PR	08-SEP-2000	2000US-02312828
PR	08-SEP-2000	2000US-02312838
PR	08-SEP-2000	2000US-02312848
PR	08-SEP-2000	2000US-02312858
PR	08-SEP-2000	2000US-02312868
PR	08-SEP-2000	2000US-02312878
PR	08-SEP-2000	2000US-02312888
PR	08-SEP-2000	2000US-02312898
PR	08-SEP-2000	2000US-02312908
PR	08-SEP-2000	2000US-02312918
PR	08-SEP-2000	2000US-02312928
PR	08-SEP-2000	2000US-02312938
PR	08-SEP-2000	2000US-02312948
PR	08-SEP-2000	2000US-02312958
PR	08-SEP-2000	2000US-02312968
PR	08-SEP-2000	2000US-02312978
PR	08-SEP-2000	2000US-02312988
PR	08-SEP-2000	2000US-02312998
PR	08-SEP-2000	2000US-02313008
PR	08-SEP-2000	2000US-02313018
PR	08-SEP-2000	2000US-02313028
PR	08-SEP-2000	2000US-02313038
PR	08-SEP-2000	2000US-02313048
PR	08-SEP-2000	2000US-02313058
PR	08-SEP-2000	2000US-02313068
PR	08-SEP-2000	2000US-02313078
PR	08-SEP-2000	2000US-02313088
PR	08-SEP-2000	2000US-02313098
PR	08-SEP-2000	2000US-02313108
PR	08-SEP-2000	2000US-02313118
PR	08-SEP-2000	2000US-02313128
PR	08-SEP-2000	2000US-02313138
PR	08-SEP-2000	2000US-02313148
PR	08-SEP-2000	2000US-02313158
PR	08-SEP-2000	2000US-02313168
PR	08-SEP-2000	2000US-02313178
PR	08-SEP-2000	2000US-02313188
PR	08-SEP-2000	2000US-02313198
PR	08-SEP-2000	2000US-02313208
PR	08-SEP-2000	2000US-02313218
PR	08-SEP-2000	2000US-02313228
PR	08-SEP-2000	2000US-02313238
PR	08-SEP-2000	2000US-02313248
PR	08-SEP-2000	2000US-02313258
PR	08-SEP-2000	2000US-02313268
PR	08-SEP-2000	2000US-02313278
PR	08-SEP-2000	2000US-02313288
PR	08-SEP-2000	2000US-02313298
PR	08-SEP-2000	2000US-02313308
PR	08-SEP-2000	2000US-02313318
PR	08-SEP-2000	2000US-02313328
PR	08-SEP-2000	2000US-02313338
PR	08-SEP-2000	2000US-02313348
PR	08-SEP-2000	2000US-02313358
PR	08-SEP-2000	2000US-02313368
PR	08-SEP-2000	2000US-02313378
PR	08-SEP-2000	2000US-02313388
PR	08-SEP-2000	2000US-02313398
PR	08-SEP-2000	2000US-02313408
PR	08-SEP-2000	2000US-02313418
PR	08-SEP-2000	2000US-02313428
PR	08-SEP-2000	2000US-02313438
PR	08-SEP-2000	2000US-02313448
PR	08-SEP-2000	2000US-02313458
PR	08-SEP-2000	2000US-02313468
PR	08-SEP-2000	2000US-02313478
PR	08-SEP-2000	2000US-02313488
PR	08-SEP-2000	2000US-02313498
PR	08-SEP-2000	2000US-02313508
PR	08-SEP-2000	2000US-02313518

PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249308P.  
PR 17-NOV-2000; 2000US-0249309P.  
PR 17-NOV-2000; 2000US-0249310P.  
PR 17-NOV-2000; 2000US-0249311P.  
PR 17-NOV-2000; 2000US-0249312P.  
PR 17-NOV-2000; 2000US-0249313P.  
PR 17-NOV-2000; 2000US-0249314P.  
PR 17-NOV-2000; 2000US-0249315P.  
PR 17-NOV-2000; 2000US-0249316P.  
PR 17-NOV-2000; 2000US-0249317P.  
PR 17-NOV-2000; 2000US-0249318P.  
PR 17-NOV-2000; 2000US-0249319P.  
PR 17-NOV-2000; 2000US-0249345P.  
PR 17-NOV-2000; 2000US-0249346P.  
PR 17-NOV-2000; 2000US-0249347P.  
PR 17-NOV-2000; 2000US-0249348P.  
PR 17-NOV-2000; 2000US-0249349P.  
PR 17-NOV-2000; 2000US-0249350P.  
PR 17-NOV-2000; 2000US-0249351P.  
PR 17-NOV-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251719P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251899P.  
PR 08-DEC-2000; 2000US-0251900P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
  
PA (HMA-) HUMAN GENOME SCI INC.  
XX  
XX  
PI Rosen CA, Baraash SC, Ruben SM;  
XX  
XX WPI; 2001-483426/52.  
XX  
XX  
XX Nucleic acids encoding human immune/haematopoietic antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating cancers and metastasis.  
XX  
XX  
XX  
PS Disclosure; SEQ ID NO 36836; 3071bp + Sequence Listing; English.  
XX  
XX  
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
CC amino acid sequences given in AAM62170 to AAM91921. (I) have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patients own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting the  
CC nucleic acid into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/haematopoietic-related diseases, especially  
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
CC to AAK67694 represent human immune/haematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
CC represent sequences used in the exemplification of the present invention  
XX  
XX Sequence 875 BP; 278 A; 210 C; 161G; 226 T; 0 U; 0 Other;



















PR	08-SEP-2000	2000US-0232080P	
PR	08-SEP-2000	2000US-0232081P	
PR	12-SEP-2000	2000US-0231568P	
PR	14-SEP-2000	2000US-0232397P	
PR	14-SEP-2000	2000US-0232398P	
PR	14-SEP-2000	2000US-0232399P	
PR	14-SEP-2000	2000US-0232400P	
PR	14-SEP-2000	2000US-0232401P	
PR	14-SEP-2000	2000US-0233063P	
PR	14-SEP-2000	2000US-0233064P	
PR	14-SEP-2000	2000US-0233065P	
PR	21-SEP-2000	2000US-0234423P	
PR	21-SEP-2000	2000US-0234474P	
PR	25-SEP-2000	2000US-0234997P	
PR	25-SEP-2000	2000US-0234998P	
PR	26-SEP-2000	2000US-0235584P	
PR	27-SEP-2000	2000US-0235583AP	
PR	27-SEP-2000	2000US-0235583BP	
PR	29-SEP-2000	2000US-0236367P	
PR	29-SEP-2000	2000US-0236368P	
PR	29-SEP-2000	2000US-0236369P	
PR	29-SEP-2000	2000US-0236370P	
PR	02-OCT-2000	2000US-0237032P	
PR	02-OCT-2000	2000US-0237033P	
PR	02-OCT-2000	2000US-0237039P	
PR	02-OCT-2000	2000US-0237040P	
PR	13-OCT-2000	2000US-0239933P	
PR	13-OCT-2000	2000US-0239937P	
PR	20-OCT-2000	2000US-0240960P	
PR	20-OCT-2000	2000US-0241212P	
PR	20-OCT-2000	2000US-0241785P	
PR	20-OCT-2000	2000US-0241876P	
PR	20-OCT-2000	2000US-0241877P	
PR	20-OCT-2000	2000US-0241809P	
PR	20-OCT-2000	2000US-0241826P	
PR	01-NOV-2000	2000US-0244617P	
PR	08-NOV-2000	2000US-0246475P	
PR	08-NOV-2000	2000US-0246476P	
PR	08-NOV-2000	2000US-0246477P	
PR	08-NOV-2000	2000US-0246478P	
PR	08-NOV-2000	2000US-0246533P	
PR	08-NOV-2000	2000US-0246534P	
PR	08-NOV-2000	2000US-0246552P	
PR	08-NOV-2000	2000US-0246556P	
PR	08-NOV-2000	2000US-0246577P	
PR	08-NOV-2000	2000US-0246582P	
PR	08-NOV-2000	2000US-0246583P	
PR	08-NOV-2000	2000US-0246609P	
PR	08-NOV-2000	2000US-0246610P	
PR	08-NOV-2000	2000US-0246611P	
PR	17-NOV-2000	2000US-0249211P	
PR	17-NOV-2000	2000US-0249212P	
PR	17-NOV-2000	2000US-0249213P	
PR	17-NOV-2000	2000US-0249214P	
PR	17-NOV-2000	2000US-0249215P	
PR	17-NOV-2000	2000US-0249216P	
PR	17-NOV-2000	2000US-0249217P	
PR	17-NOV-2000	2000US-0249218P	
PR	17-NOV-2000	2000US-0249244P	
PR	17-NOV-2000	2000US-0249245P	
PR	17-NOV-2000	2000US-0249246P	
PR	17-NOV-2000	2000US-0249255P	
PR	17-NOV-2000	2000US-0249257P	
PR	17-NOV-2000	2000US-0249259P	

PR	17-NOV-2000;	2000US-029300P.
PR	01-DEC-2000;	2000US-0250160P.
PR	01-DEC-2000;	2000US-0250391P.
PR	05-DEC-2000;	2000US-0251030P.
PR	05-DEC-2000;	2000US-0251988P.
PR	05-DEC-2000;	2000US-0256719P.
PR	06-DEC-2000;	2000US-0251479P.
PR	08-DEC-2000;	2000US-0251856P.
PR	08-DEC-2000;	2000US-0251868P.
PR	08-DEC-2000;	2000US-0251869P.
PR	08-DEC-2000;	2000US-0251989P.
PR	08-DEC-2000;	2000US-0251990P.
PR	11-DEC-2000;	2000US-0254097P.
PR	05-JAN-2001;	2001US-0259678P.
PA	(HUMA-) HUMAN GENOME SCI INC.	
XX		
PI	Rosen CA, Barash SC, Ruben SM;	
XX		
DR	WPI; 2001-465570/50.	
XX		
PT	Isolated nucleic acid molecule encoding a reproductive system antigen is used in preventing, treating or ameliorating a medical condition.	
XX		
PS	Disclosure; SEQ ID NO 5653; 1297bp + Sequence Listing; English.	
XX		
CC	The present invention provides the protein and coding sequences of a number of human reproductive system related antigens. These can be used in the prevention and treatment of reproductive system disorders, including cancer. The present sequence is a genomic sequence encoding a protein of the invention	
CC		
SQ	Sequence 1524 BP; 357 A; 369 C; 387 G; 411 T; 0 U; 0 Other;	
XX		
Query Match	21.9%; Score 65.8; DB 4; Length 1524;	
Best local similarity	68.4%; Pred. No. 1.3e-09;	
Matches	91; Conservative 0; Mismatches 42; Indels 0; Gaps 0;	
QY	89 TGATATTAGATATATATATATGTTTTCATCATAGTCTCGGTATTAACCCCTCAC 148	
DB	1086 TGCTGTGTTGTTATGATAGACAGTTTCCCTCATGATCCTGGCTCAATACTCCATAGC 1145	
QY	149 CCTTGTTACATCATTTGCTTTAAGTTGATGGTTTAGGGCTAGAAACAACCTCTCT 208	
DB	1146 CCTGTATACACTCTTTGTTGTATATTGAGCTGTGTAGGCCACAGGAACGACTCTCT 1205	
QY	209 CTCACCTCTCTC 221	
DB	1206 CCTCTCTCTGCCC 1218	
RESULT 14		
ABA07734		
ID	ABA07734 standard; DNA; 1524 BP.	
XX		
AC	ABA07734;	
XX		
DT	11-JAN-2002 (first entry)	
XX		
DE	Human ovarian and breast cancer associated polynucleotide SEQ ID NO 529.	
XX		
KW	Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;	
KW	antiallergic; hepatotrophic; antidiabetic; antiinflammatory; anticancer;	
KW	vulnerary; anticoagulant; antibacterial; antifungal; antiparasitic;	
KM	cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;	
KM	neurological disease; infection; human; secreted protein; ds.	
OS	Homo sapiens.	
XX		
PN	WO200155325-A2.	
XX		
FD	02-AUG-2001.	
XX		







PT polypeptides, useful for diagnosing, treating and/or preventing human  
PT diseases and disorders, particularly ovarian and/or breast cancer.  
XX  
PS Disclosure; SEQ ID NO 529; 577pp + Sequence Listing; English.  
XX  
CC The invention relates to novel genes (ABA07454-ABA08224) and proteins  
CC (ABA10743-ABA10980) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in  
CC the diagnosis, treatment and prevention of: (a) cancer; e.g. breast and  
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,  
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune  
CC disorders e.g. Addison's disease, allergies, autoimmune hemolytic  
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,  
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)  
CC cardiovascular disorders such as myocardial ischaemia; (d) wound healing  
CC / (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)  
CC infectious diseases such as viral, bacterial, fungal and parasitic  
CC infections. Note: The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in electronic format directly  
CC from WIPO at [http://wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences)  
XX  
SQ Sequence 1524 BP; 357 A; 369 C; 387 G; 411 T; 0 U; 0 Other;  
Query March 21.9%; Score 65.8; DB 4; Length 1524;  
Best Local Similarity 68.4%; Pred. No. 1.3e-09;  
Matches 91; Conservative 0; Mismatches 42; Indels 0; Gaps 0;  
QY 89 TGAATATTGATATATATATAGTTTGTGTCATAGTCTCGGCTATATAACCCCTCAC 148  
DB 1086 TGCTGTGTTTATGATAGACAGTTTCTCCATGATCTCTGCTATATACTCCATAGC 1145  
QY 149 CCTGTATACAGTATTTGTTATAGGTGATGCTTAAAGCCCTCAGAACAAACTCTCT 208  
DB 1146 CTTGTATACAGTCTTTTGTGTTATATGATGTGTTAGGCCACAGAAACAGACTCTCT 1205  
QY 209 CTCTCACCCTTCTC 221  
DB 1206 CTTCTCTCTGCCCC 1218  
RESULT 15  
ACH26427  
ID ACH26427 standard; cDNA; 452 BP.  
XX  
AC ACH26427;  
XX  
DT 13-OCT-2003 (first entry)  
XX  
DE Human adult ovary cDNA #4807.  
XX  
KM Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;  
KM genome mapping; biodiversity; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN US2003073623-A1.  
XX  
PD 17-APR-2003.  
XX  
PF 30-JUL-2001; 2001US-00918995.  
XX  
PR 30-JUL-2001; 2001US-00918995.  
XX  
PA (DRMA/) DRMANAC R T.  
PA (LABA/) LABAT I.  
PA (STAC/) STACHE-CRAIN B.  
PA (DICK/) DICKSON M C.  
PA (JONE/) JONES L W.  
XX  
PI Dmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;  
XX

DR WPI; 2003-615964/58.  
XX  
XX New polynucleotide sequences obtained from various cDNA libraries, useful  
PT as hybridization probes, as oligomers for PCR, for chromosome and gene  
PT mapping, in the recombinant production of protein, or in generating  
PT antisense DNA or RNA.  
XX  
PS Claim 1; SEQ ID NO 13639; 44pp; English.  
XX  
CC The invention relates to an isolated polynucleotide comprising any one of  
CC 38043 cDNA sequences, appearing as ACH12769-ACH50831, whose sequence was  
CC determined by the technique of SBH (sequencing by hybridisation). Also  
CC included is a purified polypeptide comprising a sequence corresponding to  
CC a reading frame of the novel polynucleotide. The nucleic acid sequences  
CC are useful in diagnostics as expressed sequence tags (EST) for  
CC identifying expressed genes or for physical mapping of the human genome,  
CC in forensics, in assessing biodiversity, or in identifying mutations  
CC responsible for genetic disorders and other traits. The nucleotide  
CC sequences are also useful as hybridisation probes, as oligomers for PCR,  
CC for chromosome and gene mapping, in the recombinant production of  
CC protein, or in generating antisense DNA or RNA. The purified polypeptide  
CC is useful for generating antibodies specific for it. The present sequence  
CC is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data  
CC for this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from USPTO at  
CC [seqdata.uspto.gov/sequence.html?DocID=20030073623](http://seqdata.uspto.gov/sequence.html?DocID=20030073623)  
XX  
SQ Sequence 452 BP; 113 A; 109 C; 87 G; 129 T; 0 U; 14 Other;  
Query March 21.6%; Score 65; DB 8; Length 452;  
Best Local Similarity 68.6%; Pred. No. 1.3e-09;  
Matches 120; Conservative 0; Mismatches 50; Indels 5; Gaps 2;  
QY 83 GGGTGTGATATATATATATATATAGTTTGTGTCATAGTCTCGGCTATATAACCC 142  
DB 125 GGGTCTGTGATATAGATATATATAGTTTTCATC--TATTTCTGGGCTTATACTCC 182  
QY 143 CCTACCCCTTGTACAGTCACTTTGTTATAGGTGATGTTTAAAGCCCTCAGAA---GCA 199  
DB 183 CAAAACCTTGTATAGGCTTTGTTATATGTTGGCACTTGAGGCCCTCAGAAAACAGC 242  
QY 200 AAAGTCTCTCTCACCCTTCTCCAGCCCTCTGCTCTGACCACTTCTTCC 254  
DB 243 AGGCTGTTTCTCAGATCTTCTCCGACCTCTTTCACCTGCTGCTTTCTCC 297  
Search completed: March 12, 2004, 18:07:59  
Job time : 124.53 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 12, 2004, 17:39:25 ; Search time 22.3958 Seconds

(without alignments)  
7458.540 Million cell updates/sec

Title: US-09-966-880A-10\_COPY\_6264\_6564

Perfect score: 301  
Sequence: 1 gcatcaagagacactagta.....cttcacaaagcggtcaag 301

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Issued Patents NA:\*

- 1: /cgn2\_6/prodata/2/ina/5A\_COMB.seq.\*
- 2: /cgn2\_6/prodata/2/ina/5B\_COMB.seq.\*
- 3: /cgn2\_6/prodata/2/ina/6A\_COMB.seq.\*
- 4: /cgn2\_6/prodata/2/ina/6B\_COMB.seq.\*
- 5: /cgn2\_6/prodata/2/ina/PCITUS\_COMB.seq.\*
- 6: /cgn2\_6/prodata/2/ina/backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	60.4	20.1 148567	4 US-09-801-876B-3	Sequence 3, Appl
2	60.4	20.1 148567	4 US-10-254-869-3	Sequence 3, Appl
3	55.8	18.5 174493	4 US-09-804-971A-3	Sequence 3, Appl
4	55.8	18.5 174493	4 US-10-238-709-3	Sequence 3, Appl
5	53.2	17.7 649	3 US-09-040-984-59	Sequence 59, Appl
6	53.2	17.7 649	4 US-09-123-912-59	Sequence 59, Appl
7	53.2	17.7 649	4 US-09-643-597-59	Sequence 59, Appl
8	53.2	17.7 649	4 US-09-480-884A-59	Sequence 59, Appl
9	53.2	17.7 649	4 US-09-542-615A-59	Sequence 59, Appl
10	53.2	17.7 649	4 US-09-606-421B-59	Sequence 59, Appl
11	53.2	17.7 649	4 US-09-221-107-59	Sequence 59, Appl
12	48.4	16.1 4080	4 US-09-016-434-1342	Sequence 1342, Ap
13	48.4	16.1 246240	2 US-08-724-394A-20	Sequence 20, Appl
14	48.4	16.1 246240	2 US-08-724-394A-21	Sequence 21, Appl
15	48.4	16.1 246240	2 US-08-724-394A-22	Sequence 22, Appl
16	47.6	15.8 41684	4 US-09-536-059-1	Sequence 1, Appl
17	45.8	15.2 65042	4 US-09-784-316-3	Sequence 3, Appl
18	45	15.0 3757	4 US-09-620-312D-106	Sequence 106, App
19	43.8	14.6 392000	4 US-10-027-983-11	Sequence 11, Appl
20	42	14.0 786431	4 US-09-751-389-3	Sequence 3, Appl
21	41.6	13.8 168575	4 US-09-426-290-1	Sequence 62, Appl
22	39.8	13.2 6801	4 US-10-204-708-62	Sequence 3, Appl
23	38.8	12.9 55827	4 US-09-813-133A-3	Sequence 3, Appl
24	38.6	12.8 22067	4 US-09-820-001-3	Sequence 3, Appl
25	38.2	12.7 319608	4 US-09-539-33D-1	Sequence 1, Appl
26	38.2	12.7 319608	4 US-09-679-409-1	Sequence 1, Appl
27	38	12.6 15418	4 US-09-783-203-1	Sequence 1, Appl

28	38	12.6 51552	4 US-09-733-294A-30	Sequence 30, Appl
29	37.8	12.6 98844	4 US-09-791-211-10	Sequence 10, Appl
30	37.8	12.6 392000	4 US-10-027-983-11	Sequence 11, Appl
31	37.2	12.4 73	4 US-09-621-976-7972	Sequence 7972, Ap
32	36.6	12.2 73	4 US-09-621-976-8430	Sequence 8430, Ap
33	36.6	12.2 73	4 US-09-621-976-8766	Sequence 8766, Ap
34	36.6	12.2 73	4 US-09-621-976-8783	Sequence 8783, Ap
35	36.6	12.2 523	4 US-09-621-976-1442	Sequence 1442, Ap
36	36.6	12.2 788	4 US-09-288-143-27	Sequence 27, Appl
37	36.4	12.1 262	4 US-09-621-976-189	Sequence 189, Appl
38	36.2	12.0 6801	4 US-10-204-708-61	Sequence 61, Appl
39	35.8	11.9 116592	4 US-09-818-512-3	Sequence 3, Appl
40	35	11.6 2189	3 US-08-846-020A-1	Sequence 1, Appl
41	35	11.6 2189	4 US-09-617-871-1	Sequence 1, Appl
42	34.8	11.6 7218	1 US-08-232-463-14	Sequence 14, Appl
43	34.8	11.6 31208	4 US-09-852-087-3	Sequence 3, Appl
44	34.6	11.5 388	4 US-09-621-976-18393	Sequence 18393, A
45	34.6	11.5 427	4 US-09-621-976-18392	Sequence 18392, A

## ALIGNMENTS

RESULT 1  
US-09-801-876B-3  
Sequence 3, Application US/09801876B  
Patent No. 6492155  
GENERAL INFORMATION:  
APPLICANT: YE, Jane et al  
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
FILE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
TITLE OF INVENTION: THEREOF  
FILE REFERENCE: CLO01160  
CURRENT APPLICATION NUMBER: US/09/801.876B  
CURRENT FILING DATE: 2001-03-09  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 3  
LENGTH: 148567  
TYPE: DNA  
ORGANISM: Human  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(148567)  
OTHER INFORMATION: n = A,T,C or G  
US-09-801-876B-3

Query Match 20.1% Score 60.4; DB 4; Length 148567;  
Best Local Similarity 65.7%; Pred. No. 1.4e-08;  
Matches 88; Conservative 0; Mismatches 46; Indels 0; Gaps 0;  
QY 16 TAGTAGAAGTGAAGAAGACACCAAGAGTGGAGAAACACTGCAATCATATAT 75  
DB 76269 TATCAAAAAGTGAAGTCACTCCCAAGAGATGGGAGAAATATTGCAATATATAT 76328  
QY 76 CTGATTAAGGTGTGAT 135  
DB 76329 CTACAAAGATGTGCAATTCATATATCAAAAGGGGCGGCTGTGGCTCATGCTGT 76388  
QY 136 TAAACCCCTCAACC 149  
DB 76389 AATCCAGCAATC 76402  
RESULT 2  
US-10-254-869-3  
Sequence 3, Application US/10254869  
Patent No. 6653117  
GENERAL INFORMATION:  
APPLICANT: YE, Jane et al  
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
FILE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
TITLE OF INVENTION: THEREOF



FILE REFERENCE: CU001160DIV  
CURRENT APPLICATION NUMBER: US/10/254,869  
CURRENT FILING DATE: 2002-09-26  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 3  
LENGTH: 148567  
TYPE: DNA  
ORGANISM: Human  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)...(148567)  
OTHER INFORMATION: n = A,T,C or G  
US-10-254-869-3

Query Match 20.1%; Score 60.4; DB 4; Length 148567;  
Best Local Similarity 65.7%; Pred. No. 1.4e-08;  
Matches 88; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 16 TAGTAAAGAAAGTGAAGAAAGCAACCCAGAGAGTGGAGAAACACTTGCMAATCATATAT 75  
DB 76269 TATCAAAAAGTGAAGAAAGTCAACCCAGAGATGGAGAAATATTGCAAAATCATATAT 76328  
QY 76 CTGATAGAGGTGTGATATATATATATATATAGTTTTCATAGTCTGCTGCTGCTGCTA 135  
DB 76329 CTACTAAAGATGTGATTTACATATATACAAAGGGCCAGCGCTGTGCTCATGCTGT 76388  
QY 136 TAAACCCCTCACC 149  
DB 76389 AATCCCGCAATTC 76402

RESULT 3  
US-09-804-471A-3/C  
Sequence 3, Application US/09804471A  
Patent No. 6479269  
GENERAL INFORMATION:  
APPLICANT: WEBSTER, Marion et al  
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
TITLE OF INVENTION: THEREOF  
FILE REFERENCE: CU001164  
CURRENT APPLICATION NUMBER: US/09/804,471A  
CURRENT FILING DATE: 2001-03-13  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 3  
LENGTH: 174493  
TYPE: DNA  
ORGANISM: Human  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)...(174493)  
OTHER INFORMATION: n = A,T,C or G  
US-09-804-471A-3

Query Match 18.5%; Score 55.8; DB 4; Length 174493;  
Best Local Similarity 61.0%; Pred. No. 4.5e-07;  
Matches 108; Conservative 0; Mismatches 67; Indels 2; Gaps 1;

QY 28 GAAAGACAACCCAGAGAGTGGAGAAACACTTGCMAATCATATATCTGATAGGGTT 87  
DB 69497 GAATGATACAGAAAGAGAGATGGATTAACAGAGACAGGTTGATATATATGA 69438  
QY 88 GTGATATATGATATATATATATAGTTTTCATAGTCTGCTGCTTATTAACCCCTCA 147  
DB 69437 GATATATATATATATATATATAGTTTTCATCCAGGTTCTGCTTATATATATATG 69378  
QY 148 CCCTGTACAGTATTTGTTATATAGTTTTCATAGTCTGCTGCTGCTGCTGCTGCTA 204  
DB 69377 TCCTGTACATCTTTTGTATTTAG--GGGTGCTTTAGGCTTTAGAGCAGGCTT 69323

RESULT 4  
US-10-238-709-3/C  
Sequence 3, Application US/10238709  
Patent No. 6680188  
GENERAL INFORMATION:  
APPLICANT: WEBSTER, Marion et al  
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
TITLE OF INVENTION: THEREOF  
FILE REFERENCE: CU001164DIV  
CURRENT APPLICATION NUMBER: US/10/238,709  
CURRENT FILING DATE: 2002-09-11  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 3  
LENGTH: 174493  
TYPE: DNA  
ORGANISM: Human  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)...(174493)  
OTHER INFORMATION: n = A,T,C or G  
US-10-238-709-3

Query Match 18.5%; Score 55.8; DB 4; Length 174493;  
Best Local Similarity 61.0%; Pred. No. 4.5e-07;  
Matches 108; Conservative 0; Mismatches 67; Indels 2; Gaps 1;

QY 28 GAAAGACAACCCAGAGAGTGGAGAAACACTTGCMAATCATATATCTGATAGGGTT 87  
DB 69497 GAATGATACAGAAAGAGAGATGGATTAACAGAGACAGGTTGATATATATGA 69438  
QY 88 GTGATATATGATATATATATATAGTTTTCATAGTCTGCTGCTTATTAACCCCTCA 147  
DB 69437 GATATATATATATATATATATAGTTTTCATCCAGGTTCTGCTTATATATATG 69378  
QY 148 CCCTGTACAGTATTTGTTATATAGTTTTCATAGTCTGCTGCTGCTGCTGCTGCTA 204  
DB 69377 TCCTGTACATCTTTTGTATTTAG--GGGTGCTTTAGGCTTTAGAGCAGGCTT 69323

RESULT 5  
US-09-040-984-59/C  
Sequence 59, Application US/09040984  
Patent No. 6210883  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS  
TITLE OF INVENTION: OF LUNG CANCER  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED AND BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/040,984  
FILING DATE: 18-MAR-1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: MAKI, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.456  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-622-4900



TELEFAX: 206-282-6031  
TELEX:  
INFORMATION FOR SEQ ID NO: 59:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 649 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-040-984-59

Query Match 17.7%; Score 53.2; DB 3; Length 649;  
Best Local Similarity 73.0%; Pred. No. 2.1e-07;  
Matches 81; Conservative 0; Mismatches 29; Indels 1; Gaps 1;

QY 15 TTGATGAAGAGTGAAGCAACCCACAGAGTGGAGAAACACTGGCAATCATATA 74  
DB 312 TTATCGATTAAGTGAAACCAACCCACAGAA-TGGAGAAATATTGGCAACCATGTA 254

QY 75 TCTGATTAAGGTTGTGATATTATGATATATATATAGTTTGTCCATAGT 125  
DB 253 TCTGATTAAGGTTGTGATATTATGATATATATATAGTTTGTCCATAGT 203

#### RESULT 6

US-09-123-912-59/c  
Sequence 59, Application US/09123912A  
Patent No. 6312895  
GENERAL INFORMATION:  
APPLICANT: Reed, Steven G.  
APPLICANT: Wang, Tonglong  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY OF LUNG CANCER  
FILE REFERENCE: 210121.455C1  
CURRENT APPLICATION NUMBER: US/09/123,912A  
CURRENT FILING DATE: 1998-07-27  
PRIORITY APPLICATION NUMBER: 09/040,802  
PRIORITY FILING DATE: 1998-03-18  
NUMBER OF SEQ ID NOS: 114  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 59  
LENGTH: 649  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: modified\_base  
LOCATION: (22)  
OTHER INFORMATION: Where n is a, c, g or t  
NAME/KEY: modified\_base  
LOCATION: (190)  
OTHER INFORMATION: Where n is a, c, g or t  
NAME/KEY: modified\_base  
LOCATION: (217)  
OTHER INFORMATION: Where n is a, c, g or t  
NAME/KEY: modified\_base  
LOCATION: (430)  
OTHER INFORMATION: Where n is a, c, g or t  
NAME/KEY: modified\_base  
LOCATION: (433)  
OTHER INFORMATION: Where n is a, c, g or t  
NAME/KEY: modified\_base  
LOCATION: (484)  
OTHER INFORMATION: Where n is a, c, g or t  
NAME/KEY: modified\_base  
LOCATION: (544)  
OTHER INFORMATION: Where n is a, c, g or t  
NAME/KEY: modified\_base  
LOCATION: (550)  
OTHER INFORMATION: Where n is a, c, g or t  
NAME/KEY: modified\_base  
LOCATION: (577)  
OTHER INFORMATION: Where n is a, c, g or t  
NAME/KEY: modified\_base  
LOCATION: (583)  
OTHER INFORMATION: Where n is a, c, g or t

NAME/KEY: modified\_base  
LOCATION: (594)  
OTHER INFORMATION: Where n is a, c, g or t  
US-09-123-912-59

Query Match 17.7%; Score 53.2; DB 4; Length 649;  
Best Local Similarity 73.0%; Pred. No. 2.1e-07;  
Matches 81; Conservative 0; Mismatches 29; Indels 1; Gaps 1;

QY 15 TTGATGAAGAGTGAAGCAACCCACAGAGTGGAGAAACACTGGCAATCATATA 74  
DB 312 TTATCGATTAAGTGAAACCAACCCACAGAA-TGGAGAAATATTGGCAACCATGTA 254

QY 75 TCTGATTAAGGTTGTGATATTATGATATATATATAGTTTGTCCATAGT 125  
DB 253 TCTGATTAAGGTTGTGATATTATGATATATATATAGTTTGTCCATAGT 203

#### RESULT 7

US-09-643-597-59/c  
Sequence 59, Application US/09643597  
Patent No. 6426072  
GENERAL INFORMATION:  
APPLICANT: Wang, Tonglong  
APPLICANT: Fan, Liqun  
APPLICANT: Kalos, Michael D.  
APPLICANT: Bangur, Chaitanya S.  
APPLICANT: Hosken, Nancy  
APPLICANT: Fanger, Gary R.  
APPLICANT: Li, Samuel X.  
APPLICANT: Wang, Aijun  
APPLICANT: Skeiky, Yassir A.W.  
APPLICANT: Henderson, Robert A.  
APPLICANT: McNeill, Patricia D.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
FILE REFERENCE: 210121.455C1  
CURRENT APPLICATION NUMBER: US/09/643,597  
CURRENT FILING DATE: 2000-08-21  
NUMBER OF SEQ ID NOS: 369  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 59  
LENGTH: 649  
TYPE: DNA  
ORGANISM: Homo sapien  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)-(649)  
OTHER INFORMATION: n = A,T,C or G  
US-09-643-597-59

Query Match 17.7%; Score 53.2; DB 4; Length 649;  
Best Local Similarity 73.0%; Pred. No. 2.1e-07;  
Matches 81; Conservative 0; Mismatches 29; Indels 1; Gaps 1;

QY 15 TTGATGAAGAGTGAAGCAACCCACAGAGTGGAGAAACACTGGCAATCATATA 74  
DB 312 TTATCGATTAAGTGAAACCAACCCACAGAA-TGGAGAAATATTGGCAACCATGTA 254

QY 75 TCTGATTAAGGTTGTGATATTATGATATATATATAGTTTGTCCATAGT 125  
DB 253 TCTGATTAAGGTTGTGATATTATGATATATATATAGTTTGTCCATAGT 203

#### RESULT 8

US-09-480-884A-59/c  
Sequence 59, Application US/09480884A  
Patent No. 6482597  
GENERAL INFORMATION:  
APPLICANT: Wang, Tonglong  
APPLICANT: Fan, Liqun  
APPLICANT: Hosken, Nancy A.  
APPLICANT: Kalos, Michael D.



APPLICANT: Fargert, Gary R.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY  
FILE REFERENCE: 210121.455C6  
CURRENT APPLICATION NUMBER: US/09/480,884A  
CURRENT FILING DATE: 2001-08-27  
NUMBER OF SEQ ID NOS: 330  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 59  
LENGTH: 649  
TYPE: DNA  
ORGANISM: Homo sapien  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(649)  
OTHER INFORMATION: n = A,T,C or G  
US-09-480-884A-59

Query Match 17.7%; Score 53.2; DB 4; Length 649;  
Best Local Similarity 73.0%; Pred. No. 2.1e-07;  
Matches 81; Conservative 0; Mismatches 29; Indels 1; Gaps 1;

15 TTAGTAAAGAGTGAAGAAAGCAACCCACAGAGTGGAGAGAAACACTTGCATATATATA 74  
312 TTATCGATTAAGTGAAGAAAGCAACCCACAGAG-TGGAGAGAAATATTGCCAAACCATGTA 254  
75 TCTGATTAAGGTTGTGATATTATGATATATATATAGGTTTGTCCATAGT 125  
253 TCTGATTAAGGTTGTGATATCCAGATCCAGATATATATTAAGAAATCTTACACTCAGT 203

RESULT 9  
US-09-542-615A-59/c  
Sequence 59, Application US/09542615A  
Patent No. 6518256  
GENERAL INFORMATION:  
APPLICANT: Wang, Tongtong  
APPLICANT: Fan, Liqun  
APPLICANT: Kalos, Michael D.  
APPLICANT: Bangur, Chaitanya S.  
APPLICANT: Hoeken, Nancy A.  
APPLICANT: Fargert, Gary R.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY  
FILE REFERENCE: 210121.455C8  
CURRENT APPLICATION NUMBER: US/09/542,615A  
CURRENT FILING DATE: 2000-04-14  
NUMBER OF SEQ ID NOS: 350  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 59  
LENGTH: 649  
TYPE: DNA  
ORGANISM: Homo sapien  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(649)  
OTHER INFORMATION: n = A,T,C or G  
US-09-542-615A-59

Query Match 17.7%; Score 53.2; DB 4; Length 649;  
Best Local Similarity 73.0%; Pred. No. 2.1e-07;  
Matches 81; Conservative 0; Mismatches 29; Indels 1; Gaps 1;

15 TTAGTAAAGAGTGAAGAAAGCAACCCACAGAGTGGAGAGAAACACTTGCATATATATA 74  
312 TTATCGATTAAGTGAAGAAAGCAACCCACAGAG-TGGAGAGAAATATTGCCAAACCATGTA 254  
75 TCTGATTAAGGTTGTGATATTATGATATATATAGGTTTGTCCATAGT 125  
253 TCTGATTAAGGTTGTGATATCCAGATCCAGATATATTAAGAAATCTTACACTCAGT 203

RESULT 10

US-09-606-421B-59/c  
Sequence 59, Application US/09606421B  
Patent No. 6531315  
GENERAL INFORMATION:  
APPLICANT: Wang, Tongtong  
APPLICANT: Fan, Liqun  
APPLICANT: Kalos, Michael D.  
APPLICANT: Bangur, Chaitanya S.  
APPLICANT: Hoeken, Nancy  
APPLICANT: Fargert, Gary R.  
APPLICANT: Li, Samuel X.  
APPLICANT: Skeiky, Yasir A.W.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
FILE REFERENCE: 210121.455C9  
CURRENT APPLICATION NUMBER: US/09/606,421B  
CURRENT FILING DATE: 2000-06-28  
NUMBER OF SEQ ID NOS: 358  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 59  
LENGTH: 649  
TYPE: DNA  
ORGANISM: Homo sapien  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(649)  
OTHER INFORMATION: n = A,T,C or G  
US-09-606-421B-59

Query Match 17.7%; Score 53.2; DB 4; Length 649;  
Best Local Similarity 73.0%; Pred. No. 2.1e-07;  
Matches 81; Conservative 0; Mismatches 29; Indels 1; Gaps 1;

15 TTAGTAAAGAGTGAAGAAAGCAACCCACAGAGTGGAGAGAAACACTTGCATATATATA 74  
312 TTATCGATTAAGTGAAGAAAGCAACCCACAGAG-TGGAGAGAAATATTGCCAAACCATGTA 254  
75 TCTGATTAAGGTTGTGATATTATGATATATATAGGTTTGTCCATAGT 125  
253 TCTGATTAAGGTTGTGATATCCAGATCCAGATATATTAAGAAATCTTACACTCAGT 203

RESULT 11  
US-09-221-107-59/c  
Sequence 59, Application US/09221107  
Patent No. 6660838  
GENERAL INFORMATION:  
APPLICANT: Wang, Tongtong  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY OF LUNG CANCER  
FILE REFERENCE: 210121.455C2  
CURRENT APPLICATION NUMBER: US/09/221,107  
CURRENT FILING DATE: 1998-12-22  
NUMBER OF SEQ ID NOS: 161  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 59  
LENGTH: 649  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: modified\_base  
LOCATION: (22)  
OTHER INFORMATION: Where n is a, c, g or t  
FEATURE:  
NAME/KEY: modified\_base  
LOCATION: (190)  
OTHER INFORMATION: Where n is a, c, g or t  
FEATURE:  
NAME/KEY: modified\_base  
LOCATION: (217)  
OTHER INFORMATION: Where n is a, c, g or t  
FEATURE:  
NAME/KEY: modified\_base

15 TTAGTAAAGAGTGAAGAAAGCAACCCACAGAGTGGAGAGAAACACTTGCATATATATA 74  
312 TTATCGATTAAGTGAAGAAAGCAACCCACAGAG-TGGAGAGAAATATTGCCAAACCATGTA 254  
75 TCTGATTAAGGTTGTGATATTATGATATATATAGGTTTGTCCATAGT 125  
253 TCTGATTAAGGTTGTGATATCCAGATCCAGATATATTAAGAAATCTTACACTCAGT 203



```
LOCATION: (430)
OTHER INFORMATION: where n is a, c, g or t
FEATURE:
NAME/KEY: modified_base
LOCATION: (433)
OTHER INFORMATION: where n is a, c, g or t
FEATURE:
NAME/KEY: modified_base
LOCATION: (484)
OTHER INFORMATION: where n is a, c, g or t
FEATURE:
NAME/KEY: modified_base
LOCATION: (544)
OTHER INFORMATION: where n is a, c, g or t
FEATURE:
NAME/KEY: modified_base
LOCATION: (550)
OTHER INFORMATION: where n is a, c, g or t
FEATURE:
NAME/KEY: modified_base
LOCATION: (577)
OTHER INFORMATION: where n is a, c, g or t
FEATURE:
NAME/KEY: modified_base
LOCATION: (583)
OTHER INFORMATION: where n is a, c, g or t
FEATURE:
NAME/KEY: modified_base
LOCATION: (594)
OTHER INFORMATION: where n is a, c, g or t
US-09-221-107-59

Query Match 17.7%; Score 53.2; DB 4; Length 649;
Best Local Similarity 73.0%; Pred. No. 2.1e-07;
Matches 81; Conservative 0; Mismatches 29; Indels 1; Gaps 1;

Dn 15 TTACTAGAAAGTAAAGCAACCCACAGAGTGGAGAAAACCTTGCAATCATAT 74
312 TTATCGATTAAGTAAAGCAACCCACAGAGTGGAGAAAATATTGGAAACCATGTA 254

Qy 75 TCTGATTAAGGTTGTGATATATATATATATATATATATATATATATATAT 125
253 TCTGATTAAGGTTGTGATATATATATATATATATATATATATATATATATAT 203

Db 253 TCTGATTAAGGTTGTGATATATATATATATATATATATATATATATATATAT 203

RESULT 12
US-09-016-434-1342
Sequence 1342, Application US/09016434
Patent No. 6500938
GENERAL INFORMATION:
APPLICANT: Janice Au-Young
APPLICANT: Jeffrey J. Sellhamer
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
TITLE OF INVENTION: PATHWAY GENE EXPRESSION
NUMBER OF SEQUENCES: 1490
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/016,434
FILING DATE: HEREWITH
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
```

```
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Zeller, Karen J.
REGISTRATION NUMBER: 37,071
REFERENCE/DOCKET NUMBER: PA-0002 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 845-4166
TELEFAX: (650) 845-0555
INFORMATION FOR SEQ ID NO: 1342:
SEQUENCE CHARACTERISTICS:
LENGTH: 4080 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GENBANK
CLONE: 934764
US-09-016-434-1342

Query Match 16.1%; Score 48.4; DB 4; Length 4080;
Best Local Similarity 70.9%; Pred. No. 1.7e-05;
Matches 78; Conservative 0; Mismatches 31; Indels 1; Gaps 1;

Dn 14 CTTAGTAAGAAAGTAAAGCAACCCACAGAGTGGAGAAAACCTTGCAATCATAT 73
3159 CCTATCAAGAAAGTAAAGCAACCCACAGAGTGGAGAAAATATTGGAGATCATAT 3217

Qy 74 ATCTGATTAAGGTTGTGATATATATATATATATATATATATATATATATAT 123
3218 ATATGATTAAGGTTGTGATATATATATATATATATATATATATATATATATAT 3267

Db 3218 ATATGATTAAGGTTGTGATATATATATATATATATATATATATATATATATAT 3267

RESULT 13
US-08-724-394A-20/c
Sequence 20, Application US/08724394A
Patent No. 5872237
GENERAL INFORMATION:
APPLICANT: Feder, John N.
APPLICANT: Krommal, Gregory S.
APPLICANT: Laney, Peter M.
APPLICANT: Ruddy, David A.
APPLICANT: Thomas, Winston
APPLICANT: Tsuchihashi, Zenta
APPLICANT: Wolfe, Roger K.
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237e1
TITLE OF INVENTION: Sequences and Antibodies Thereeto
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,394A
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Pites, Renee A.
REGISTRATION NUMBER: 35,136
REFERENCE/DOCKET NUMBER: 017957-000100
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-576-0200
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
```



LENGTH: 246240 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 1..246240  
OTHER INFORMATION: /note= "HLA-H.CONTIG"  
US-08-724-394A-20

Query Match 16.1%; Score 48.4; DB 2; Length 246240;  
Best Local Similarity 77.8%; Pred. No. 0.00012;  
Matches 84; Conservative 0; Mismatches 21; Indels 3; Gaps 2;

QY 186 AGGCTCAGAGCAAACTCTCTCTCAGCCCTCTGCTCTGCGACCTC 245  
DB 173083 AGACCGCAGAGCCAAAGTTTCTCTGATCTCTCCTGCCCTCTCTGAGTCC-C 173025  
QY 246 ATTCTTCCCTGAGGCCA--CATAGAACTAGAACTCTCTTCCACAG 291  
DB 173024 ATTCTCCCCGAGCTAGCCATAGAAATGGAATTCCTTCTCTCAAG 172977

RESULT 14  
US-08-724-394A-21/C

/ Sequence 21, Application US/08724394A

/ Patent No. 5872237

/ GENERAL INFORMATION:

/ APPLICANT: Feder, John N.

/ APPLICANT: Krommal, Gregory S.

/ APPLICANT: Lauer, Peter M.

/ APPLICANT: Ruddy, David A.

/ APPLICANT: Thomas, Winston

/ APPLICANT: Teuchhashi, Zenta

/ APPLICANT: Wolff, Roger K.

/ TITLE OF INVENTION: Megabase Transcript Map: No. 5872237e1

/ TITLE OF INVENTION: Sequences and Antibodies Thereo

/ NUMBER OF SEQUENCES: 31

/ CORRESPONDENCE ADDRESSES:

/ ADDRESSES: TOWNSEND and TOWNSEND and CREM LLP

/ STREET: Two Embarcadero Center, 8th Floor

/ CITY: San Francisco

/ STATE: CA

/ COUNTRY: USA

/ ZIP: 94111-3834

/ COMPUTER READABLE FORM:

/ MEDIUM TYPE: Floppy disk

/ COMPUTER: IBM PC compatible

/ OPERATING SYSTEM: PC-DOS/MS-DOS

/ SOFTWARE: Patentin Release #1.0, Version #1.30

/ CURRENT APPLICATION DATA:

/ APPLICATION NUMBER: US/08/724,394A

/ FILING DATE: 01-OCT-1996

/ CLASSIFICATION: 536

/ ATTORNEY/AGENT INFORMATION:

/ NAME: Fitts, Renee A.

/ REGISTRATION NUMBER: 35,136

/ REFERENCE/DOCKET NUMBER: 017957-000100

/ TELECOMMUNICATION INFORMATION:

/ TELEPHONE: 415-576-0200

/ TELEFAX: 415-576-0300

/ INFORMATION FOR SEQ ID NO: 21:

/ SEQUENCE CHARACTERISTICS:

/ LENGTH: 246240 base pairs

/ TYPE: nucleic acid

/ STRANDEDNESS: not relevant

/ TOPOLOGY: not relevant

/ MOLECULE TYPE: cDNA

/ FEATURE:

/ NAME/KEY: misc feature

/ LOCATION: 1..246240

/ OTHER INFORMATION: /note= "HLA-H.CONTIG"

US-08-724-394A-21

Query Match 16.1%; Score 48.4; DB 2; Length 246240;  
Best Local Similarity 77.8%; Pred. No. 0.00012;  
Matches 84; Conservative 0; Mismatches 21; Indels 3; Gaps 2;

QY 186 AGGCTCAGAGCAAACTCTCTCTCAGCCCTCTGCTCTGCGACCTC 245  
DB 173083 AGACCGCAGAGCCAAAGTTTCTCTGATCTCTCCTGCCCTCTCTGAGTCC-C 173025  
QY 246 ATTCTTCCCTGAGGCCA--CATAGAACTAGAACTCTCTTCCACAG 291  
DB 173024 ATTCTCCCCGAGCTAGCCATAGAAATGGAATTCCTTCTCTCAAG 172977

RESULT 15

US-08-724-394A-22/C

/ Sequence 22, Application US/08724394A

/ Patent No. 5872237

/ GENERAL INFORMATION:

/ APPLICANT: Feder, John N.

/ APPLICANT: Krommal, Gregory S.

/ APPLICANT: Lauer, Peter M.

/ APPLICANT: Ruddy, David A.

/ APPLICANT: Thomas, Winston

/ APPLICANT: Teuchhashi, Zenta

/ APPLICANT: Wolff, Roger K.

/ TITLE OF INVENTION: Megabase Transcript Map: No. 5872237e1

/ TITLE OF INVENTION: Sequences and Antibodies Thereo

/ NUMBER OF SEQUENCES: 31

/ CORRESPONDENCE ADDRESSES:

/ ADDRESSES: TOWNSEND and TOWNSEND and CREM LLP

/ STREET: Two Embarcadero Center, 8th Floor

/ CITY: San Francisco

/ STATE: CA

/ COUNTRY: USA

/ ZIP: 94111-3834

/ COMPUTER READABLE FORM:

/ MEDIUM TYPE: Floppy disk

/ COMPUTER: IBM PC compatible

/ OPERATING SYSTEM: PC-DOS/MS-DOS

/ SOFTWARE: Patentin Release #1.0, Version #1.30

/ CURRENT APPLICATION DATA:

/ APPLICATION NUMBER: US/08/724,394A

/ FILING DATE: 01-OCT-1996

/ CLASSIFICATION: 536

/ ATTORNEY/AGENT INFORMATION:

/ NAME: Fitts, Renee A.

/ REGISTRATION NUMBER: 35,136

/ REFERENCE/DOCKET NUMBER: 017957-000100

/ TELECOMMUNICATION INFORMATION:

/ TELEPHONE: 415-576-0200

/ TELEFAX: 415-576-0300

/ INFORMATION FOR SEQ ID NO: 22:

/ SEQUENCE CHARACTERISTICS:

/ LENGTH: 246240 base pairs

/ TYPE: nucleic acid

/ STRANDEDNESS: not relevant

/ TOPOLOGY: not relevant

/ MOLECULE TYPE: cDNA

/ FEATURE:

/ NAME/KEY: misc feature

/ LOCATION: 1..246240

/ OTHER INFORMATION: /note= "HLA-H.CONTIG"

US-08-724-394A-22

Query Match 16.1%; Score 48.4; DB 2; Length 246240;  
Best Local Similarity 77.8%; Pred. No. 0.00012;  
Matches 84; Conservative 0; Mismatches 21; Indels 3; Gaps 2;

QY 186 AGGCTCAGAGCAAACTCTCTCTCAGCCCTCTGCTCTGCGACCTC 245  
DB 173083 AGACCGCAGAGCCAAAGTTTCTCTGATCTCTCCTGCCCTCTCTGAGTCC-C 173025



Oy 246 ATTCTTCCCTGAGGCA--CATAGAACTAGATCTCTCTTCCACAAG 291  
Db 173024 ATTCTCCCGGAGTCTAGCCATAGAAATGAGAAATTCCTCTCTCAAG 172977

Search completed: March 13, 2004, 00:50:58  
Job time : 27.3958 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM nucleic - nucleic search, using sw model

Run on: March 12, 2004, 21:22:51 / Search time 105.627 Seconds

(without alignments)  
10491.276 Million cell updates/sec

Title: US-09-966-880a-10\_COPY\_6264\_6564

Perfect score: 301 1 gacatcaagagacacttagta.....cttcacaaagggcgctcaag 301

Sequence:

Scoring table: IDENTITY\_NUC

Searched: 243257 seqs, 184079884 residues

Total number of hits satisfying chosen parameters: 4865114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

Published Applications NA:\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
17: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
18: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	301	100.0	6564	9	US-09-966-880a-10
2	77.6	25.8	511	15	US-10-027-632-92073
3	77.6	25.8	511	15	US-10-027-632-317886
4	74.4	24.7	3298	15	US-10-027-632-112733
5	74.4	24.7	3298	15	US-10-027-632-112734
6	74.4	24.7	164875	15	US-10-085-117-332
7	73	24.3	367378	14	US-10-312-841-1
8	67	22.3	35465	14	US-10-161-572-6
9	67	22.3	35465	14	US-10-161-572-6
10	66.6	22.1	568	15	US-10-027-632-49688
11	66.6	22.1	1155	15	US-10-027-632-117444
12	66.6	22.1	1770	15	US-10-027-632-248967
13	66	21.9	431	12	US-10-085-783A-24956
14	66	21.9	431	15	US-10-242-535A-24956
15	65.8	21.9	1524	10	US-09-764-891-5553

16	65.8	21.9	1524	14	US-10-205-428-529	Sequence 529, App
17	65.6	21.8	521	15	US-10-027-632-62410	Sequence 62410, A
18	65.6	21.8	521	15	US-10-027-632-63014	Sequence 63014, A
19	65.6	21.8	521	15	US-10-027-632-306536	Sequence 306536, A
20	65.2	21.7	593	15	US-10-027-632-287770	Sequence 287770, A
21	65	21.6	452	10	US-09-918-995-13639	Sequence 13639, A
22	64.4	21.4	13069	9	US-09-764-869-1850	Sequence 1850, Ap
23	64.4	21.4	13069	14	US-10-091-504-1850	Sequence 1850, Ap
24	64.4	21.4	13069	15	US-10-027-577-1850	Sequence 1850, Ap
25	64.4	21.4	32203	15	US-09-764-869-1849	Sequence 1849, Ap
26	64.4	21.4	32203	14	US-10-091-504-1849	Sequence 1849, Ap
27	64.4	21.4	32203	15	US-10-027-577-1849	Sequence 1849, Ap
28	62.6	20.8	1691139	14	US-10-067-514-1	Sequence 1, Appl
29	62.6	20.8	1691139	15	US-10-419-723-1	Sequence 1, Appl
30	62.4	20.7	531	14	US-10-029-386-10489	Sequence 10489, A
31	62.4	20.7	1480	14	US-10-017-161-337	Sequence 337, App
32	62.4	20.7	1480	15	US-10-027-798-301	Sequence 301, App
33	61.4	20.4	315	12	US-10-085-783A-30007	Sequence 30007, A
34	61.4	20.4	315	15	US-10-242-535A-30007	Sequence 30007, A
35	60.4	20.1	148567	9	US-09-801-876B-3	Sequence 3, Appl
36	60.4	20.1	148567	12	US-10-667-442-3	Sequence 3, Appl
37	60.4	20.1	148567	14	US-10-254-869-3	Sequence 3, Appl
38	59.4	19.7	604	15	US-10-027-632-275507	Sequence 275507, A
39	59	19.6	604	15	US-10-027-632-275506	Sequence 275506, A
40	58.2	19.3	39776	14	US-10-160-293-3	Sequence 3, Appl
41	57.4	19.1	416	10	US-09-918-995-36136	Sequence 36136, A
42	57.2	19.0	761	15	US-10-027-632-15636	Sequence 15636, A
43	57	18.9	160361	12	US-10-235-192A-35	Sequence 35, Appl
44	56.6	18.8	5879	14	US-10-311-455-241	Sequence 241, App
45	56.4	18.7	3673778	14	US-10-312-841-2	Sequence 2, Appl

## ALIGNMENTS

RESULT 1	US-09-966-880a-10	US-10-027-632-92073
1	Sequence 10, Appl	Sequence 92073, A
2	Patent No. US20020164743A1	Sequence 92073, A
3	GENERAL INFORMATION:	Sequence 92073, A
4	APPLICANT: Muramatsu, Masamichi	Sequence 92073, A
5	TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE	Sequence 92073, A
6	FILE REFERENCE: 06501-088001	Sequence 92073, A
7	CURRENT APPLICATION NUMBER: US/09/966, 880A	Sequence 92073, A
8	CURRENT FILING DATE: 2001-09-28	Sequence 92073, A
9	PRIOR APPLICATION NUMBER: PCT/JP00/01918	Sequence 92073, A
10	PRIOR FILING DATE: 2000-03-28	Sequence 92073, A
11	PRIOR APPLICATION NUMBER: JP 11-371382	Sequence 92073, A
12	PRIOR FILING DATE: 1999-12-27	Sequence 92073, A
13	PRIOR APPLICATION NUMBER: JP 11-178999	Sequence 92073, A
14	PRIOR FILING DATE: 1999-06-24	Sequence 92073, A
15	PRIOR APPLICATION NUMBER: JP 11-87192	Sequence 92073, A
16	PRIOR FILING DATE: 1999-03-29	Sequence 92073, A
17	NUMBER OF SEQ ID NOS: 36	Sequence 92073, A
18	SOFTWARE: FASTSEQ for Windows Version 4.0	Sequence 92073, A
19	SEQ ID NO 10	Sequence 92073, A
20	LENGTH: 6564	Sequence 92073, A
21	TYPE: DNA	Sequence 92073, A
22	ORGANISM: Homo sapiens	Sequence 92073, A
23	US-09-966-880a-10	Sequence 92073, A
24	Query Match	100.0%; Score 301; DB 9; Length 6564;
25	Best Local Similarity	100.0%; Pred. No. 3.4e-81; Indels 0; Gaps 0;
26	Matches 301; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
27	1 GCATCAAGGACCTTAGTAGAAGAGTGAAGCAACCCACAGAGTGGAGAGAAACAC 60	
28	6264 GCATCAAGGACCTTAGTAGAAGAGTGAAGCAACCCACAGAGTGGAGAGAAACAC 6263	
29	61 TTGCAATCATATATCTGATTAAGGCTTGTGATATTAATATATATAGATTGTTTC 120	
30	6324 TTGCAATCATATATCTGATTAAGGCTTGTGATATTAATATATATAGATTGTTTC 6383	



QY 121 ATAGTTCCTGGCTTAAACCCCTCACCTTGTAGCATTTGTATAAGGTTGAT 180  
Db 6384 ATAGTTCCTGGCTTAAACCCCTCACCTTGTAGCATTTGTATAAGGTTGAT 6443  
QY 181 GGTTAGGCTCAGAGCAAACTCTCTCTCACTTCTCCAGCCCTCTGTCTGGC 240  
Db 6444 GGTTAGGCTCAGAGCAAACTCTCTCTCACTTCTCCAGCCCTCTGTCTGGC 6503  
QY 241 ACCCTATTCTTCCCTGAGGCCACATAGAACTAGATCTCTTCCACAAGGCGTCAA 300  
Db 6504 ACCCTATTCTTCCCTGAGGCCACATAGAACTAGATCTCTTCCACAAGGCGTCAA 6563  
QY 301 G 301  
Db 6564 G 6564

## RESULT 2

US-10-027-632-92073/c  
; Sequence 92073, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 92073  
; LENGTH: 511  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-92073

Query Match 25.8%; Score 77.6; DB 15; Length 511;  
Best Local Similarity 68.6%; Pred. No. 2,2e-13;  
Matches 107; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 66 AATCATATATCTGATAGAGGTTGTGATATATATATATATAGTTTGTCCATAGT 125  
Db 463 ACTATTAAGCTTACTTGGCAGTGTGATATATATATATATAGTTTGTCCAGT 404  
QY 126 TCTGTGCTATTAACCCCTCACCTTGTAGCATTTGTATAGTGGATGTTT 185  
Db 403 TCTGTGATCATACTCCATAGCCCTTGTAGCATTTTTCATATATGTTGGTGT 344  
QY 186 AGGCTCAGAGCAAACTCTCTCTCACTTCTC 221  
Db 343 AGGCTCAGAGCAAGGCTCTTACTTTCTCTGCCC 308

RESULT 3  
US-10-027-632-317886/c  
; Sequence 317886, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:

; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 317886  
; LENGTH: 511  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-317886

Query Match 25.8%; Score 77.6; DB 15; Length 511;  
Best Local Similarity 68.6%; Pred. No. 2,2e-13;  
Matches 107; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 66 AATCATATATCTGATAGAGGTTGTGATATATATATATATAGTTTGTCCATAGT 125  
Db 463 ACTATTAAGCTTACTTGGCAGTGTGATATATATATATATAGTTTGTCCAGT 404  
QY 126 TCTGTGCTATTAACCCCTCACCTTGTAGCATTTGTATAGTGGATGTTT 185  
Db 403 TCTGTGATCATACTCCATAGCCCTTGTAGCATTTTTCATATATGTTGGTGT 344  
QY 186 AGGCTCAGAGCAAACTCTCTCTCACTTCTC 221  
Db 343 AGGCTCAGAGCAAGGCTCTTACTTTCTCTGCCC 308

## RESULT 4

US-10-027-632-112733/c  
; Sequence 112733, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:

; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 112733







QY	Db	QY	Db
1093010	ATTTCACACCTTAAATAATCTATACGATATTAATTTATATATAAATTTTATCCATAT	1093210	ATTTCACACCTTAAATAATCTATACGATATTAATTTATATATAAATTTTATCCATAT
1093015	TCTCGGCTTATAAACCCCTCACCCTTGTTGTCAGTCATTTGTTATTAAGTTGAGTGGTT	1093215	TCTCGGCTTATAAACCCCTCACCCTTGTTGTCAGTCATTTGTTATTAAGTTGAGTGGTT
1093020	TCCCTACCTTATTAATTCCTCATACCTTATTCGCATCTATTTATTAATAATTAATACCTT	1093220	TCCCTACCTTATTAATTCCTCATACCTTATTCGCATCTATTTATTAATAATTAATACCTT
1093025	AGGCTCAGAAAGCAAACTC-----TCTCTCGACCTCTCCAGCGCCCTCTC	1093225	AGGCTCAGAAAGCAAACTC-----TCTCTCGACCTCTCCAGCGCCCTCTC
1093030	AAACCTCAAAAACAATCTCAATTAAGCAAAATCTCTCTAACCTTCTCTTACCCCTCATTT	1093230	AAACCTCAAAAACAATCTCAATTAAGCAAAATCTCTCTAACCTTCTCTTACCCCTCATTT
1093035	TCTGGACCTCATTTCTTCCCTGAGGCGCATAGAAACTAGATCTCTCT	1093235	TCTGGACCTCATTTCTTCCCTGAGGCGCATAGAAACTAGATCTCTCT
1093040	ATCTACTCTTATCTCTCCAAAACAAAATCTTTACCCATCTTTCTAT	1093240	ATCTACTCTTATCTCTCCAAAACAAAATCTTTACCCATCTTTCTAT

```

RESULT 8
US-10-161-572-6/c
; Sequence 6, Application US/10161572
; Publication No. US20030087266A1
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: IGS AS MODIFIERS OF THE p53 PATHWAY AND METHODS OF USE
; FILE REFERENCE: EX02-097C-PC
; CURRENT APPLICATION NUMBER: US/10/161,572
; PRIOR FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: US 60/296,076
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/328,605
; PRIOR FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 60/338,733
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 60/357,253
; PRIOR FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: US 60/357,600
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 35465
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-161-572-6

```

	Query Match	Similarity	22.3%	Score 67	DB 14	Length 35465
	Best Local	Similarity	72.7%	Pred. No. 3,6e-09		
	Matches	113	Conservative	0	Mismatches 45	Indels 9
					Gaps	1
QY	85	GTGTGATATATATGATATATATAGTGTGTTCATAGTTCCTGAGCTTAATAACCC	144			
DB	4807	GTGGATATATATATATATATATATGTTTCATCCAGTTCCTGGGTGTAATCCTTA	4748			
QY	145	TCACCCCTGTTACACGTCAATTTGTTATAAGTTGATGGTTTAGGCC-----TCAGA	195			
DB	4747	CAGCCTTTGTACACGTCTCTGTTTATATGTTGTGTACCTTAAGGCTGAGGACAGGTCTGA	4688			
QY	196	AGCAAAATCTCTCTCTCACTTTCCAGCCCTCCTGTCTCGGAC	242			
DB	4687	AGAAACAAATCTCTCTGACCTCTCCCTCCGCCCTCATTTCTCGCCCC	4641			

```

RESULT 9
US-10-161-572-8/C
; Sequence 8, Application US/10161572
; Publication No. US2003087266A1
; GENERAL INFORMATION:
; APPLICANT: EXELIXIS, INC.
; TITLE OF INVENTION: IGS AS MODIFIERS OF THE p53 PATHWAY AND METHODS OF USE
; FILE REFERENCE: EX02-097C-PC
; CURRENT APPLICATION NUMBER: US/10/161,572
; CURRENT FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: US 60/296,076
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/328,605

```

```

; PRIOR FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: US 60/338,723
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 60/357,253
; PRIOR FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: US 60/357,600
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 36891
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-161-572-8

```

Query March	22.3%	Score 67	DB 14	Length 36991
Best Local Similarity	67.7%	Pred. No. 3,7e-09		
Matches 113	Conservative 0	Mismatches 45	Indels 9	Gaps 1

  

QY	85	GTGTGATATTATGATATATATAGGTTTGTGCATAGTCTGTGACTATTAACCCCC	144
DB	4807	GTGTATTATATATATATAATATATGTTTCATCCACAGTCTGTGCTGTAACCTCTA	4748
QY	145	TCACCCCTTTACAGCTATTGTTATAGGTGATAGGTTTAAGCC-----TCGA	195
DB	4747	CAGCCTTTGTACAGCTCTCTGTTATATATGTTGTAACCTTAAAGCTGGGGCAGGTCTGA	4688
QY	196	AGCAAAATCTCTCTTCACCTTTCCAGCCCTCTGTCTGTGCAC	242
DB	4687	AGAAACGAATCTCTTCAGCCTCTCTGCCCCCTTCATTCTGTGCC	4661

```

RESULT 10
US-10-027-632-49688/c
; Sequence 49688, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Mann, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 49688
; LENGTH: 568
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-49688

```

	Query Match	Similarity	Score	DB	Length
Best Local	67.9%	Pred. 66.5-101	568		
Matches	93	Conservative	0	Mismatches	44
				Indels	0
				Gaps	0
QY	85	GTGTGATATATGATATATATATATGTTTGTTCATAGTTCTCGGCTTATTAACCCC	144		
DB	332	GTGCGTATATATGATATATATATGTTTTCATCATGTTCTCGGCTCAAACTCGA	263		







Search completed: March 13, 2004, 05:30:45  
Job time : 116.627 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: March 12, 2004, 12:27:59 ; Search time 273.899 Seconds  
(without alignments)  
13767.261 Million cell updates/sec

Title: US-09-966-880a-11  
Perfect score: 87  
Sequence: 1 agagaccatcatatcga.....ctgacaccactatgacag 87

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

GeneBdb1:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_scs:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: gb\_vl:\*  
16: em\_ba:\*  
17: em\_fun:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_scs:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rnd:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vtc:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	87	100.0	87	BD016836	BD016836 Novel cyt
2	87	100.0	2818	BD016833	BD016833 Novel cyt
3	87	100.0	5514	BD016834	BD016834 Novel cyt
4	87	100.0	11204	BD016860	BD016860 Novel cyt
5	87	100.0	11204	AB040430	AB040430 Homo sapi
6	87	100.0	71132	AC092184	AC092184 Homo sapi
7	84	96.6	1828	BC006286	BC006286 Homo sapi
8	84	96.6	2791	AB040431	AB040431 Homo sapi
9	41	1767	10	AB091291	AB091291 Mus muscu
10	41	1767	10	BD016828	BD016828 Novel cyt
11	41	2440	6	AF132979	AF132979 Mus muscu
12	41	47.1	178130	AC119975	AC119975 Mus muscu
13	34.6	39.8	241757	AC094826	AC094826 Rattus no
14	34.6	39.8	259506	AC109119	AC109119 Rattus no
15	34.6	39.8	345098	AC120617	AC120617 Rattus no
16	32.6	37.5	228932	AL045669	AL045669 Mouse DNA
17	32.6	37.5	231634	AC101661	AC101661 Mus muscu
18	32.2	37.0	175970	AC097488	AC097488 Homo sapi
19	31.4	36.1	63101	AP003489	AP003489 Oryza sat
20	31.4	36.1	157739	AC122485	AC122485 Mus muscu
21	31.4	36.1	160545	AP003508	AP003508 Oryza sat
22	31.2	35.9	134137	HS3B5	AL022239 Human DNA
23	31.2	35.9	165953	AC129243	AC129243 Rattus no
24	31.2	35.9	288358	AC106121	AC106121 Rattus no
25	31.1	35.6	137520	EC045985	AL459965 Equus cab
26	30.6	34.5	155735	CNS01RGY	AL160234 Human chr
27	30	34.5	265272	AC132111	AC132111 Mus muscu
28	29.8	34.3	184037	AC146658	AC146658 Oryza sat
29	29.8	34.3	193410	AC020594	AC020594 Homo sapi
30	29.6	34.0	179124	AL037170	AL037170 Mouse DNA
31	29.6	34.0	237087	AC128203	AC128203 Rattus no
32	29.6	34.0	247899	AC025909	AC025909 Mus muscu
33	29.6	34.0	249713	AC108233	AC108233 Rattus no
34	29.6	34.0	260894	AC119873	AC119873 Mus muscu
35	29.6	34.0	261378	AC122982	AC122982 Rattus no
36	29.4	33.8	185694	AC079505	AC079505 Mus muscu
37	29.4	33.8	213674	AC079528	AC079528 Mus muscu
38	29.4	33.8	216330	AC098443	AC098443 Rattus no
39	29.2	33.6	123325	AP006116	AP006116 Lotus cor
40	29.2	33.6	152422	AC016491	AC016491 Homo sapi
41	29.2	33.6	164519	AC020630	AC020630 Homo sapi
42	29.2	33.6	215353	AC137981	AC137981 Mus muscu
43	29.2	33.6	231074	AC142509	AC142509 Rattus no
44	29.2	33.6	231140	AC097799	AC097799 Rattus no
45	29	33.3	76836	AC119618	AC119618 Homo sapi

## ALIGNMENTS

RESULT 1  
LOCUS BD016836  
DEFINITION Novel cytidine deaminase.  
ACCESSION BD016836  
VERSION BD016836.1 GI:22558012  
KEYWORDS JP 2001245669-A/9.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 87)  
AUTHORS Honjo, T. and Muramatsu, M.  
TITLE Novel cytidine deaminase  
JOURNAL Patent: JP 2001245669-A 9 11-SEP-2001;



COMMENT JAPAN TOBACCO INC,TASUKU HONJO  
OS Homo sapiens (human)  
PN JP 2001245669-A/9  
PD 11-SEP-2001  
PF 28-MAR-2000 JP 2000092981  
PI TASUKU HONJO,MASAMICHI MURAMATSU  
PC C12N15/09,A61K39/395,A61P11/06,A61P13/12,  
PC A61P17/00,  
PC A61P27/02,A61P37/02,A61P37/08,C07K16/18,C12N1/19, PC  
C12N1/21,  
PC C12N5/10,C12N9/78,C12P21/02,C12P21/08/(C12N1/21,C12R1:19), PC  
(C12N5/10,C12R1:91),C12N15/00,C12N5/00,C12N5/00,C12R1:91) CC  
FH Key Location/Qualifiers  
1..87  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

FEATURES  
source  
1..87  
Location/Qualifiers

ORIGIN  
Query Match 100.0%; Score 87; DB 6; Length 87;  
Best Local Similarity 100.0%; Pred. No. 1e-20;  
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAAGACCATCATTAATTGAGTGAATTTCTGGCCCTGAGACTTGAGGAGGCAAGA 60  
DB 1 AGAAGACCATCATTAATTGAGTGAATTTCTGGCCCTGAGACTTGAGGAGGCAAGA 60

QY 61 AGACACTCTGGACACCATATGAGCAG 87  
DB 61 AGACACTCTGGACACCATATGAGCAG 87

RESULT 2  
BD016833  
LOCUS BD016833 2818 bp DNA linear PAT 27-AUG-2002  
DEFINITION Novel cytidine deaminase.  
ACCESSION BD016833.1 GI:22558009  
VERSION JP 2001245669-A/6.  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 2818)  
AUTHORS Honjo,T. and Muramatsu,M.  
TITLE Novel cytidine deaminase  
JOURNAL Patent: JP 2001245669-A 6 11-SEP-2001;  
JAPAN TOBACCO INC,TASUKU HONJO  
OS Homo sapiens (human)  
PN JP 2001245669-A/6  
PD 11-SEP-2001  
PF 28-MAR-2000 JP 2000092981  
PI TASUKU HONJO,MASAMICHI MURAMATSU  
PC C12N15/09,A61K39/395,A61P11/06,A61P13/12,  
PC A61P17/00,  
PC A61P27/02,A61P37/02,A61P37/08,C07K16/18,C12N1/19, PC  
C12N1/21,  
PC C12N5/10,C12N9/78,C12P21/02,C12P21/08/(C12N1/21,C12R1:19), PC  
(C12N5/10,C12R1:91),C12N15/00,C12N5/00,C12N5/00,C12R1:91) CC  
FH Key Location/Qualifiers  
1..87  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

FEATURES  
source  
1..2818  
Location/Qualifiers

ORIGIN  
Query Match 100.0%; Score 87; DB 6; Length 2818;  
Best Local Similarity 100.0%; Pred. No. 1.2e-20;

Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAAGACCATCATTAATTGAGTGAATTTCTGGCCCTGAGACTTGAGGAGGCAAGA 60  
DB 1 AGAAGACCATCATTAATTGAGTGAATTTCTGGCCCTGAGACTTGAGGAGGCAAGA 60

QY 61 AGACACTCTGGACACCATATGAGCAG 87  
DB 61 AGACACTCTGGACACCATATGAGCAG 87

RESULT 3  
BD016834  
LOCUS BD016834 5514 bp DNA linear PAT 27-AUG-2002  
DEFINITION Novel cytidine deaminase.  
ACCESSION BD016834.1 GI:22558010  
VERSION JP 2001245669-A/7.  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 5514)  
AUTHORS Honjo,T. and Muramatsu,M.  
TITLE Novel cytidine deaminase  
JOURNAL Patent: JP 2001245669-A 7 11-SEP-2001;  
JAPAN TOBACCO INC,TASUKU HONJO  
OS Homo sapiens (human)  
PN JP 2001245669-A/7  
PD 11-SEP-2001  
PF 28-MAR-2000 JP 2000092981  
PI TASUKU HONJO,MASAMICHI MURAMATSU  
PC C12N15/09,A61K39/395,A61P11/06,A61P13/12,  
PC A61P17/00,  
PC A61P27/02,A61P37/02,A61P37/08,C07K16/18,C12N1/19, PC  
C12N1/21,  
PC C12N5/10,C12N9/78,C12P21/02,C12P21/08/(C12N1/21,C12R1:19), PC  
(C12N5/10,C12R1:91),C12N15/00,C12N5/00,C12N5/00,C12R1:91) CC  
FH Key Location/Qualifiers  
1..5514  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

FEATURES  
source  
1..5514  
Location/Qualifiers

ORIGIN  
Query Match 100.0%; Score 87; DB 6; Length 5514;  
Best Local Similarity 100.0%; Pred. No. 1.3e-20;  
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAAGACCATCATTAATTGAGTGAATTTCTGGCCCTGAGACTTGAGGAGGCAAGA 60  
DB 1032 AGAAGACCATCATTAATTGAGTGAATTTCTGGCCCTGAGACTTGAGGAGGCAAGA 1091

QY 61 AGACACTCTGGACACCATATGAGCAG 87  
DB 1092 AGACACTCTGGACACCATATGAGCAG 1118

RESULT 4  
BD016860  
LOCUS BD016860 11204 bp DNA linear PAT 27-AUG-2002  
DEFINITION Novel cytidine deaminase.  
ACCESSION BD016860.1 GI:22558036  
VERSION JP 2001245669-A/33.  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.



REFERENCE 1 (bases 1 to 11204)  
 AUTHORS Honjo, T. and Muramatsu, M.  
 TITLE Novel cytidine deaminase  
 JOURNAL Patent: JP 2001245668-A 33 11-SEP-2001;  
 JAPAN TOBACCO INC./TASUKU HONJO  
 COMMENT OS Homo sapiens (human)  
 PN JP 2001245669-A/33  
 PD 11-SEP-2001  
 PF 26-MAR-2000 JP 2000092981  
 PI TASUKU HONJO, MASAMICHI MORIMATSU  
 PC C12N15/09, A61K39/395, A61P1/00, A61P11/06, A61P13/12,  
 PC A61P17/00,  
 PC A61P27/02, A61P27/16, A61P37/02, A61P37/08, C07K16/18, C12N1/19, PC  
 C12N1/21,  
 PC C12N5/10, C12N9/78, C12P21/02, C12P21/08//C12N1/21, C12R1/19, PC  
 (C12N5/10, C12R1/91), C12N15/00, C12N5/00, (C12N5/00, C12R1/91) CC  
 FH Key  
 FEATURES Location/Qualifiers  
 source 1..11204  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 ORIGIN  
 Query Match 100.0%; Score 87; DB 6; Length 11204;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-20;  
 Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 AGAAGACCATCATTAATTAAGTGAATTTTCTGGCCCTGAGACTTGAGGAGGCAAGA 60  
 Db 442 AGAAGACCATCATTAATTAAGTGAATTTTCTGGCCCTGAGACTTGAGGAGGCAAGA 501  
 Oy 61 AGAAGACTTGAGACCACTATGAGACAG 87  
 Db 502 AGAAGACTTGAGACCACTATGAGACAG 528  
 RESULT 5  
 AB040430 11204 bp DNA linear PRI 03-OCT-2000  
 LOCUS Homo sapiens AID gene for activation-induced cytidine deaminase,  
 DEFINITION complete cds.  
 ACCESSION AB040430  
 VERSION AB040430.1 GI:9988407  
 KEYWORDS AID; activation-induced cytidine deaminase.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (sites)  
 AUTHORS Muto, T., Muramatsu, M., Tanigaki, M., Kinoshita, K. and Honjo, T.  
 TITLE Isolation, tissue distribution, and chromosomal localization of the  
 human activation-induced cytidine deaminase (AID) gene  
 JOURNAL Genomics 68 (1), 85-88 (2000)  
 MEDLINE 20408890  
 PUBMED 10950930  
 REFERENCE 2 (sites)  
 AUTHORS Revy, P., Muto, T., Levy, Y., Geissmann, F., Plehani, A., Sanal, O.,  
 Catalan, N., Forveille, M., Dufourcq-Lageat, R., Gennery, A.,  
 Tzschann, I., Ercsey, F., Kayserili, H., Ugazio, A. G., Brousse, N.,  
 Muramatsu, M., Notarangelo, L. D., Kinoshita, K., Honjo, T., Fischer, A.  
 and Durandy, A.  
 TITLE Activation-induced cytidine deaminase (AID) deficiency causes the  
 autosomal recessive form of the Hyper-IgM syndrome (HIGM2)  
 JOURNAL Cell 102 (5), 565-575 (2000)  
 MEDLINE 20460541  
 PUBMED 11007475  
 REFERENCE 3 (bases 1 to 11204)  
 AUTHORS Muto, T., Muramatsu, M., Tanigaki, M., Kinoshita, K. and Honjo, T.  
 TITLE Direct Substitution  
 Submitted (18-MAR-2000) Tasuku Honjo, Kyoto University, Department  
 of Medical Chemistry, Faculty of Medicine, Yoshida, Sakyo-ku,  
 Kyoto, Kyoto 606-8501, Japan (E-mail: honjo@four.med.kyoto-u.ac.jp).

FEATURES  
 source  
 Location/Qualifiers  
 1..11204  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 gene join(521..528,6280..6427,7807..8077,8371..8486,8956..9009)  
 /gene="AID"  
 join(521..528,6280..6427,7807..8077,8371..8486,8956..9009)  
 /gene="AID"  
 /codon\_start=1  
 /product="activation-induced cytidine deaminase"  
 /protein\_id="BAB12720.1"  
 /db\_xref="GI:9988408"  
 /translation="MDSLNNRKKFYGRKNRMKGRRETYCYVVRKRDSTFSL  
 DFGYLRKNGCHVELFLRYISDMWDHPGRCTRVTFWFSMPCYDCAHVADFLRNP  
 MSLRIFLALYPCEDRKAEPEGLRLHRAQVIALMTKDYFCWNTVEHNERFK  
 AMEGLHNSVRLSRQRLRLLPLVEVDIDRDAFRITGL"  
 ORIGIN  
 Query Match 100.0%; Score 87; DB 6; Length 11204;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-20;  
 Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 AGAAGACCATCATTAATTAAGTGAATTTTCTGGCCCTGAGACTTGAGGAGGCAAGA 60  
 Db 442 AGAAGACCATCATTAATTAAGTGAATTTTCTGGCCCTGAGACTTGAGGAGGCAAGA 501  
 Oy 61 AGAAGACTTGAGACCACTATGAGACAG 87  
 Db 502 AGAAGACTTGAGACCACTATGAGACAG 528  
 RESULT 6  
 AC092184 71132 bp DNA linear PRI 12-JUN-2002  
 LOCUS Homo sapiens 12 BAC RP11-43817 (Roswell Park Cancer Institute Human  
 BAC library) complete sequence.  
 ACCESSION AC092184 AC013443  
 VERSION AC092184.7 GI:21206067  
 KEYWORDS HNG.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 71132)  
 AUTHORS Wuzny, D. M., Adams, C., Adio-Oduola, B., Allousman, F. R., Allen, C.,  
 Alsbrooks, S. L., Amaralunga, H. C., Are, J. R., Ayale, M., Banks, T.,  
 Barbarella, J., Benton, J., Bimaga, K., Blankenburg, K., Bonnin, D.,  
 Bouck, J., Bowie, S., Brileva, M., Brown, E., Brown, M., Bryant, N. P.,  
 Buhray, C., Burch, P., Burkett, C., Burrell, K. L., Byrd, N. C.,  
 Carron, T. F., Carter, M., Cavazos, S. R., Checko, J., Chavez, D.,  
 Chen, G., Chen, R., Chen, Z., Chiu, D., Chowdhry, I., Christopoulos, C.,  
 Cleveland, C. D., Cox, C., Coyle, M. D., Dabrone, S. R., David, R.,  
 Davila, M. L., Davis, C., Davy-Carroll, L., Dederich, D. A.,  
 Delaney, K. R., Delgado, O., Denn, A. L., Ding, Y., Dinh, H. H.,  
 Douthwaite, K. J., Draper, R., Dugan-Rocha, S., Durbin, K. J.,  
 Eberhart, C., Edgar, D., Edwards, C. C., Elhaj, C., Emswiler, S.,  
 Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P.,  
 Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N.,  
 Gill, R., Gorrell, J. H., Guevara, W., Gunaratne, P., Hale, S.,  
 Hamilton, K., Han, J., Harris, C., Harris, K., Hart, M., Havlak, P.,  
 Hawes, A., Hernandez, J., Hernandez, O., Hodson, A., Hogue, M.,  
 Holloway, C., Hollins, B., Homel, P., Howard, S., Huber, J., Hulik, S.,  
 Hume, J., Ioshikhes, I., Jackson, J. E., Jacobson, B., Jia, Y.,  
 Johnson, R., Jolivet, S., Joudan, S., Karlsson, S., Kelly, S., Khan, U.,  
 King, L., Kovach, J., Kovar, C., Kratochvil, J., Kureishi, A., Landry, N.,  
 Leal, B., Lee, E., Lewis, L. C., Lewis, L. J., Li, J., Li, Z., Lichtarge, O.,  
 Lieu, C., Liu, J., Liu, W., Lousseng, H., Lozano, R. J., Lu, X.,  
 Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P.,  
 Marondel, I., Martin, R., Martindale, A., Martinez, B., Maesey, B.,  
 Mawhinney, E., McLeod, M. P., Meador, M., Mei, G., Messcher, S.,  
 Metzker, M., Miller, A., Miner, G., Miner, Z., Mitchell, T.,



standards - estimated error rate less than 1 per 10,000 bases.  
 Reports of lowest quality individual bases and measures of base  
 quality are listed below. Description of the metrics can be found  
 at URL:  
<http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

#### QUALSTAT-REPORT

##### Location/Qualifiers

```

1..71132
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-438L7"
/complement(1..1448)
/feature="overlaps bases 1..1448 of clone AC092490"
/function="clone overlap"
187..408
/standard_name="57233"
439..560
/standard_name="92005"
774..881
/rpt_family="MIR"
903..1190
/rpt_family="AluSg"
1191..1213
/rpt_family="AT_rich"
1591..1807
/standard_name="6612"
1744..1819
/standard_name="8198"
1966..2264
/rpt_family="AluSg"
3296..3328
/rpt_family="TTTC)n"
4097..4249
/rpt_family="AluSg"
4360..4654
/rpt_family="AluY"
complement(5140..5262)
/rpt_family="FLAM_C"
5275..5304
/rpt_family="AT_rich"
5305..5563
/rpt_family="AluJo"
5564..5599
/rpt_family="AT_rich"
5644..5733
/rpt_family="MSTD"
5734..6012
/rpt_family="AluSg"
6020..6336
/rpt_family="AluSx"
6337..6629
/rpt_family="MSTD"
6630..6650
/rpt_family="CAA)n"
complement(6711..6817)
/rpt_family="MIR"
complement(7033..7327)
/rpt_family="AluSx"
7607..7745
/rpt_family="FLAM_A"
7980..8047
/rpt_family="MIR"
complement(8050..8549)
/rpt_family="LTR47A"
8567..8698
/rpt_family="MIR"
complement(8815..9101)
/rpt_family="AluSg"
complement(11227..12175)
/rpt_family="LTR5"

```

#### TITLE

Unpublished  
 2 (bases 1 to 71132)

#### REFERENCE

Worley, K.C.  
 Direct Submission  
 Submitted (25-JUN-2001) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

#### REFERENCE

Worley, K.C.

Direct Submission  
 Submitted (18-MAY-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

#### REFERENCE

Worley, K.C.

Direct Submission  
 Submitted (25-MAY-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

#### REFERENCE

Worley, K.C.

Direct Submission  
 Submitted (12-JUN-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA

#### COMMENT

On May 25, 2002 this sequence version replaced gi:20901754.  
 INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email  
[gc-help@bcm.tmc.edu](mailto:gc-help@bcm.tmc.edu)

CLONE LENGTH: This sequence does not necessarily represent the  
 entire insert of this clone. Overlapping regions of clones are only  
 sequenced and submitted once, so the sequence for the remainder of  
 the insert may be found in the record for the adjacent clones.  
 Overlapping clones are noted at the beginning and end of the  
 features listing.

#### ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches  
 of a local database that includes entries from dbSTS, GDB, and  
 local mapping efforts.  
 Repeats are identified using RepeatMasker (A. Smit and P. Green,  
 unpublished.) for Human and Mouse sequences.  
 Genes and Region of sequence similarity are identified by BLAST  
 (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the  
 EST and cDNA sequences. Genes demonstrate at least two exons  
 flanked by consensus splice sites that maintained sequence  
 continuity across the splice junctions. Sequences that are not  
 identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum  
 standard of double strand coverage with a minimum of 2 clones and 2  
 reads with no ambiguities or 2 chemistries with a minimum of 2  
 clones and 3 reads with no ambiguities. If the sequence quality for  
 a region does not meet this standard, it will be indicated in the  
 annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality



STS 12625..12900  
 /standard\_name="87433"  
 repeat\_region complement(13547..13664)  
 /rpt\_family="L1MB5"  
 repeat\_region complement(13665..13979)  
 /rpt\_family="L1MB5"  
 repeat\_region complement(13980..14193)  
 /rpt\_family="L1MB5"  
 repeat\_region 14632..14924  
 /rpt\_family="L1MB5"  
 repeat\_region complement(15813..17136)  
 /rpt\_family="L1MB5"  
 repeat\_region complement(17272..17395)  
 /rpt\_family="L1MB5"  
 repeat\_region 17426..17575  
 /rpt\_family="L1MB5"

Query Match 100.0%; Score 87; DB 9; Length 71132;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-20;  
 Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAGACCATCATTAATGAAGTGAATTTCTGGCCTGAGACTTGACGGAGGCAAGA 60  
 DB 34990 AGAGACCATCATTAATGAAGTGAATTTCTGGCCTGAGACTTGACGGAGGCAAGA 35049  
 QY 61 AGACACTCTGGACACCACTATGACAG 87  
 DB 35050 AGACACTCTGGACACCACTATGACAG 35076

RESULT 7  
 LOCUS BC006296 1828 bp mRNA linear PRI 03-OCT-2003  
 DEFINITION Homo sapiens activation-induced cytidine deaminase, mRNA (CDNA clone MGC:12911 IMAGE:4054915), complete cds.  
 ACCESSION BC006296  
 VERSION BC006296.2 GI:33871601  
 KEYWORDS MGC.  
 ORGANISM Homo sapiens (human)  
 SOURCE Homo sapiens  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 1 (bases 1 to 1828)  
 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hale, F., Dietzenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, K.B., Bonaldo, M.F., Casavant, T.L., Scheet, T.E., Brownstein, M.J., Udell, T.B., Toshiyuki, S., Abramson, R.D., Mallah, S.J., Bosak, S.A., McEwan, P.J., Abramson, R.D., Mallah, S.J., Bosak, S.A., McEwan, P.J., Wotley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hui, S.W., Villalón, D.K., Mizny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Rodriguez, S., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, J.S., Krzywicki, M.I., Skalska, U., Smalls, D.E., Schermer, A., Schein, J.E., Jones, S.J., and Marra, M.A.  
 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

JOURNAL MEDLINE 22388257  
 PUBMED 12477932  
 REFERENCES 2 (bases 1 to 1828)  
 STRAUSBERG, R.  
 Direct Submission  
 Submitted (09-APR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

## REMARK

NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 On Aug 19, 2003 this sequence version replaced gi:13623400.  
 Contact: MGC help desk  
 Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
 Tissue Procurement: Louis Staudt  
 CDNA Library Preparation: Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),  
 Gaithersburg, Maryland;  
 Web site: <http://www.nisc.nih.gov/>  
 Contact: [nisc.mgc@nih.gov](mailto:nisc.mgc@nih.gov)  
 Akhter, N., Ayale, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Green, K., Brinkley, C., Brooks, S., Dietrich, N.L., Grant, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Larc, P., Legaspi, R., Maduro, Q.L., Masello, C., Maskell, B., Mastrian, S.D., McCloskey, D.C., McDowell, J., Pearson, R., Santipol, S., Thomas, P.J., Touchman, J.W., Tsurgan, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>  
 Series: IRAL Plate: 17 Row: a Column: 1  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 10190699.  
 Location/Qualifiers  
 1..1828  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="MGC:12911 IMAGE:4054915"  
 /cvs\_type="Primary B-Cells from Tonsils"  
 /clone\_lib="NIH MGC 46"  
 /lab\_host="DH10B-R"  
 /note="Vector: pOTB7"  
 1..1828  
 /gene="AICDA"  
 /note="synonyms: AID, HIGM2, CD42, ARP2"  
 /db\_xref="LocusID:57379"  
 /db\_xref="MIM:605257"  
 77..673  
 /codon\_start=1  
 /product="activation-induced cytidine deaminase"  
 /protein\_id="AAH06296.1"  
 /db\_xref="GI:13623401"  
 /db\_xref="LocusID:57379"  
 /translation="MDSILNMRKELPQFKNVRAKGRRETYLCVVRSDSFTSL DPGYLRNKGCHVEELFARYISDNLDGRCRYRVTPSNKSCYCAHVAADFRGNP NLSLITFARLFCEDRAEPGLRLRRAAGQIIMPFKDFYCMNTEVEHETFK AAEGLHENSVRSLRRLRILPLVEVDLDRFRLIG"

## FEATURES

source  
 1..1828  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="MGC:12911 IMAGE:4054915"  
 /cvs\_type="Primary B-Cells from Tonsils"  
 /clone\_lib="NIH MGC 46"  
 /lab\_host="DH10B-R"  
 /note="Vector: pOTB7"  
 1..1828  
 /gene="AICDA"  
 /note="synonyms: AID, HIGM2, CD42, ARP2"  
 /db\_xref="LocusID:57379"  
 /db\_xref="MIM:605257"  
 77..673  
 /codon\_start=1  
 /product="activation-induced cytidine deaminase"  
 /protein\_id="AAH06296.1"  
 /db\_xref="GI:13623401"  
 /db\_xref="LocusID:57379"  
 /translation="MDSILNMRKELPQFKNVRAKGRRETYLCVVRSDSFTSL DPGYLRNKGCHVEELFARYISDNLDGRCRYRVTPSNKSCYCAHVAADFRGNP NLSLITFARLFCEDRAEPGLRLRRAAGQIIMPFKDFYCMNTEVEHETFK AAEGLHENSVRSLRRLRILPLVEVDLDRFRLIG"

## gene

1..1828  
 /gene="AICDA"  
 /note="synonyms: AID, HIGM2, CD42, ARP2"  
 /db\_xref="LocusID:57379"  
 /db\_xref="MIM:605257"  
 77..673  
 /codon\_start=1  
 /product="activation-induced cytidine deaminase"  
 /protein\_id="AAH06296.1"  
 /db\_xref="GI:13623401"  
 /db\_xref="LocusID:57379"  
 /translation="MDSILNMRKELPQFKNVRAKGRRETYLCVVRSDSFTSL DPGYLRNKGCHVEELFARYISDNLDGRCRYRVTPSNKSCYCAHVAADFRGNP NLSLITFARLFCEDRAEPGLRLRRAAGQIIMPFKDFYCMNTEVEHETFK AAEGLHENSVRSLRRLRILPLVEVDLDRFRLIG"

## CDS

1..1828  
 /product="activation-induced cytidine deaminase"  
 /protein\_id="AAH06296.1"  
 /db\_xref="GI:13623401"  
 /db\_xref="LocusID:57379"  
 /translation="MDSILNMRKELPQFKNVRAKGRRETYLCVVRSDSFTSL DPGYLRNKGCHVEELFARYISDNLDGRCRYRVTPSNKSCYCAHVAADFRGNP NLSLITFARLFCEDRAEPGLRLRRAAGQIIMPFKDFYCMNTEVEHETFK AAEGLHENSVRSLRRLRILPLVEVDLDRFRLIG"

## ORIGIN

Query Match 96.6%; Score 84; DB 9; Length 1828;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-19;  
 Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GAACATCATTAATGAAGTGAATTTCTGGCCTGAGACTTGACGGAGGCAAGA 63  
 DB 1 GAACATCATTAATGAAGTGAATTTCTGGCCTGAGACTTGACGGAGGCAAGA 60  
 QY 64 CACTTGACACCACTATGACAG 87  
 DB 61 CACTTGACACCACTATGACAG 84

RESULT 8  
 LOCUS AB040431 2791 bp mRNA linear PRI 03-OCT-2000  
 DEFINITION Homo sapiens AID mRNA for activation-induced cytidine deaminase, complete cds.  
 ACCESSION AB040431  
 VERSION AB040431.1 GI:9988409



KEYWORDS AID; activation-induced cytidine deaminase; Human AID.  
 SOURCE Homo sapiens (human)  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (sites) Muto, T., Muramatsu, M., Taniguchi, M., Kinoshita, K. and Honjo, T. Isolation, tissue distribution, and chromosomal localization of the human activation-induced cytidine deaminase (AID) gene  
 Genomics 68 (1), 85-88 (2000)  
 JOURNAL MEDLINE 20408850  
 PUBMED 10950930  
 REFERENCE 2 (sites) Revy, P., Muto, T., Levy, Y., Gelsmann, F., Plebani, A., Sanal, O., Catalan, N., Forville, M., Dufourcq-Lageat, R., Gennery, A., Tezcan, I., Ersoy, F., Kayseril, H., Ugazio, A.G., Brouse, N., Muramatsu, M., Notarangelo, L.D., Kinoshita, K., Honjo, T., Fischer, A. and Durandy, A. Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2)  
 Cell 102 (5), 565-575 (2000)  
 JOURNAL MEDLINE 20460541  
 PUBMED 11007475  
 REFERENCE 3 (bases 1 to 2791) Muto, T., Muramatsu, M., Taniguchi, M., Kinoshita, K. and Honjo, T. Direct Submission  
 Submitted (18-MAR-2000) Tsukuba Honjo, Kyoto University, Department of Medical Chemistry, Faculty of Medicine, Yoshida, Sakyo-ku, Kyoto, Kyoto 606-8501, Japan (E-mail: honjo@imf.med.kyoto-u.ac.jp, Tel:81-75-753-4371(ex.4371), Fax:81-75-753-4388)  
 FEATURES  
 Location/Qualifiers  
 1..2791  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 1..2791  
 /gene="AID"  
 77..673  
 /gene="AID"  
 /codon\_start=1  
 /product="activation-induced cytidine deaminase"  
 /protein\_id="BAB12721.1"  
 /db\_xref="GI:9988410"  
 /translation="MDSILMRKFLYQFKNRKAKGRREYLCVYVRDASPSFLDFGLKNGCHEVLLFLYISDWLDPERCVRVTFWFSFPCDCAHVADPFRGNDNSLRFTFLYFCEDRKAEPLGLRLHAGVQALIMFCKYFCWNTFVNHRTKXAWEGHENSVRSLRQLRILPLVEVDLRDAFRLGL"  
 ORIGIN  
 Query Match 96.6%; Score 84; DB 9; Length 2791;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-19;  
 Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 4 GAACATCATTAATGAGTATTTCTGCGCTGAGACTTGAGGAGGACGAAGA 63  
 DB 1 GAACATCATTAATGAGTATTTCTGCGCTGAGACTTGAGGAGGACGAAGA 60  
 QY 64 CACTTGACACCACTATGAGACG 87  
 DB 61 CACTTGACACCACTATGAGACG 84  
 RESULT 9  
 AB091291 1767 bp DNA linear ROD 19-SEP-2003  
 LOCUS flanking region and partial cde.  
 DEFINITION  
 ACCESSION AB091291  
 VERSION AB091291.1 GI:34447115  
 KEYWORDS Mus musculus (house mouse)  
 SOURCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 Gonda, H., Sugai, M., Nambu, Y., Katakai, T., Agata, Y., Yokota, Y. and Shimizu, A.  
 Id2: Inhibitor of Pax5 and E2A in B cell activation  
 Unpublished  
 2 (bases 1 to 1767)  
 Gonda, H., Sugai, M., Nambu, Y., Katakai, T., Agata, Y., Yokota, Y. and Shimizu, A.  
 Direct Submission  
 Submitted (04-SEP-2002) Hiroyuki Gonda, Center for Molecular Biology and Genetics; 53 Shogoin-Tawahara-cho, Sakyo-ku, Kyoto 606-8507, Japan (E-mail: hgonda@evirus.kyoto-u.ac.jp, Tel:81-75-751-4189, Fax:81-75-751-4190)  
 FEATURES  
 Location/Qualifiers  
 1..1767  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="6"  
 <1..1759  
 /note="5' flanking region"  
 1760..1767  
 /gene="AID"  
 1760..1767  
 /gene="AID"  
 /codon\_start=1  
 /product="activation-induced cytidine deaminase"  
 /protein\_id="BAC84974.1"  
 /db\_xref="GI:34447116"  
 /db\_xref="LOCUSID:11628"  
 /translation="MD"  
 ORIGIN  
 Query Match 47.1%; Score 41; DB 10; Length 1767;  
 Best Local Similarity 75.3%; Pred. No. 0.00075;  
 Matches 64; Conservative 0; Mismatches 20; Indels 1; Gaps 1;  
 QY 4 GAACATCATTAATGAGTATTTCTGCGCTGAGACTTGAGGAGGACGAAGA 62  
 DB 1683 GAACATCATTAATGAGTATTTCTGCGCTGAGACTTGAGGAGGACGAAGA 1742  
 QY 63 ACACTTGACACCACTATGAGACG 87  
 DB 1743 TCACTTGACACCACTATGAGACG 1767  
 RESULT 10  
 BD016828 2440 bp DNA linear PAT 27-AUG-2002  
 LOCUS Novel cytidine deaminase.  
 DEFINITION  
 ACCESSION BD016828  
 VERSION BD016828.1 GI:22558004  
 KEYWORDS Novel cytidine deaminase  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 2440)  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 Honjo, T. and Muramatsu, M.  
 Novel cytidine deaminase  
 Patent: JP 2001245669-A 1 11-SEP-2001;  
 JOURNAL JAPAN TOBACCO INC, TASUKU HONJO  
 COMMENT  
 OS Mus musculus (mouse)  
 PN JP 2001245669-A/1  
 PD 11-SEP-2001  
 PF 28-MAR-2000 JP 2000092381  
 PI TASUKU HONJO, MASAMICHI MURAMATSU  
 PC C12N15/09, A61K39/395, A61K39/395, A61P1/00, A61P11/06, A61P13/12, A61P17/00,  
 PC A61P27/02, A61P37/02, A61P37/08, C07K16/18, C12N1/19, PC C12N1/21,  
 PC C12N5/10, C12N9/78, C12P21/02, C12P21/08//C12N1/21, C12N1/19, PC



(C12N5/10,C12R1:91),C12N5/00,C12N5/00,(C12N5/00,C12R1:91) CC  
FH Key Location/Qualifiers  
FT CDS (93) . (689)  
FT 5'UTR (1) . (192)  
FT 3'UTR (680) . (2440).

FEATURES  
source  
1. .2440  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"

## ORIGIN

Query Match 47.1%; Score 41; DB 6; Length 2440;  
Best Local Similarity 75.3%; Pred. No. 0.00076;  
Matches 64; Conservative 0; Mismatches 20; Indels 1; Gaps 1;

QY 4 GAACCATCTATTAAATGAGATTCTTGCCCTGAGACTTGACGAGGACAG-AGA 62  
DB 16 GAAGCAGCCCTTGCTTGAAGCAAGCTTCTTGCCCTGAAGACTTGAGGAGTCAAGAAAG 75  
QY 63 ACACCTGTGACACCACTATGACAG 87  
DB 76 TCACGCTGAGACCGATATGACAG 100

RESULT 11  
AF132979 2440 bp mRNA linear ROD 05-AUG-1999  
LOCUS AF132979  
DEFINITION Mus musculus activation-induced cytidine deaminase (AID) mRNA,  
complete cds.

ACCESSION AF132979  
VERSION AF132979.1 GI:5305727  
KEYWORDS  
SOURCE Mus musculus  
ORGANISM Mus musculus (house mouse)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS  
TITLE  
1 (bases 1 to 2440)  
Muramatsu, M., Sankaranarayanan, V.S., Anant, S., Sugai, M., Kinoshita, K.,  
Davidson, N.O. and Honjo, T.  
Specific expression of activation-induced cytidine deaminase (AID),  
a novel member of the RNA-editing deaminase family in germinal  
center B cells

JOURNAL J. Biol. Chem. 274 (26), 18470-18476 (1999)  
MEDLINE 99303612  
PubMed 10373455  
REFERENCE 2 (bases 1 to 2440)  
AUTHORS Muramatsu, M. and Honjo, T.  
TITLE Direct Substitution  
Submitted (04-MAR-1999) Medical Chemistry, Kyoto University, Konoe  
Yoshida Sakyo-ku, Kyoto 606-8501, Japan  
JOURNAL Location/Qualifiers  
1. .2440  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"

gene  
1. .2440  
/gene="AID"  
93. .689  
/gene="AID"  
/note="APOBEC-1 homologue"  
/codon\_start=1  
/product="activation-induced cytidine deaminase"  
/protein\_id="AAD41793.1"

CDS  
/db\_xref="GI:5305728"  
/translation="MDSLMKQKPKLYHKRNRMAKGRHETLYCVVRRPSATSCSL  
DSGLNKGSCGCHVELFLRYISDWDDPGRCTRVTFWSMPCICAHVAEFLKWP  
NSLRFTLRALFYCEDRKAPGELRLHLAGVQIGIMTFKDFYQWNTFVNRRTFK  
AMEGLHNSVRLTRQLRRLPLLYEVDLRDAFRLMGF"

ORIGIN  
Query Match 47.1%; Score 41; DB 10; Length 2440;  
Best Local Similarity 75.3%; Pred. No. 0.00076;

Matches 64; Conservative 0; Mismatches 20; Indels 1; Gaps 1;

QY 4 GAACCATCTATTAAATGAGATTCTTGCCCTGAGACTTGACGAGGACAG-AGA 62  
DB 16 GAAGCAGCCCTTGCTTGAAGCAAGCTTCTTGCCCTGAAGACTTGAGGAGTCAAGAAAG 75

QY 63 ACACCTGTGACACCACTATGACAG 87  
DB 76 TCACGCTGAGACCGATATGACAG 100

RESULT 12  
AC119975 178130 bp DNA linear HTG 18-MAR-2003  
LOCUS AC119975  
DEFINITION Mus musculus clone RP24-483K3, WORKING DRAFT SEQUENCE, 8 unordered  
pieces.

ACCESSION AC119975  
VERSION AC119975.3 GI:28951252  
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE  
AUTHORS  
TITLE  
JOURNAL Unpublished  
2 (bases 1 to 178130)  
Birtten, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,  
Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,  
Bouhagalter, B., Brown, A., Camarero, J., Campopiano, A., Chang, D.,  
Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,  
Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,  
Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,  
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
Kamat, A., Karakas, A., Kells, C., Lacroque, K., Lamazares, R.,  
Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G.,  
Maclean, C., MacDonald, P., Major, J., Margulis, N., Matthews, C.,  
McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menues, L.,  
Mhova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,  
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,  
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,  
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,  
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schnupack, R.,  
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
Topham, K., Travers, M., Travis, N., Triggilo, J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, M., Young, G.,  
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

JOURNAL Direct Submission  
Submitted (02-MAY-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 178130)  
REFERENCE  
AUTHORS

Birtten, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,  
Anderson, S., Arachchi, H.M., Barna, N., Baerlein, V., Bloom, T.,  
Boguslavsky, L., Bouhagalter, B., Camarero, J., Chang, J., Choepel, Y.,  
Collymore, A., Cooke, P., Corum, B., Dearellano, K.,  
Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Faro, S.,  
Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S.,  
Graham, L., Grand-Pierre, N., Hale, N., Hagopian, D., Hagos, B.,  
Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
Kamat, A., Karakas, A., Kells, C., Landers, T., Levine, R.,  
Lindblad-Toh, K., Liu, G., Liu, A., Mabbitt, R., Maclean, C.,  
MacDonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,  
Meldrum, J., Menues, L., Mhova, T., Mlenga, V., Murphy, T., Naylor, J.,  
Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P.,  
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,  
Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,  
Roman, J., Schauer, S., Schnupack, R., Seaman, S., Severy, P., Smith, C.,  
Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,  
Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M.,  
Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X.,















TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

Unpublished  
2 (bases 1 to 345098)  
Worley, K.C.  
Direct Submission  
Submitted (09-May-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 345098)  
Rat Genome Sequencing Consortium.  
Direct Submission  
Submitted (15-Nov-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Nov 15, 2002 this sequence version replaced gi:23265477.  
The sequence in this assembly is a combination of BAC based reads  
and whole genome shotgun sequencing reads assembled using Atlas  
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described  
in the feature table below represents a scaffold in the Atlas  
assembly (a 'contig-scaffold'). Within each contig-scaffold,  
individual sequence contigs are ordered and oriented, and separated  
by sized gaps filled with Ns to the estimated size. The sequence  
may extend beyond the ends of the clone and there may be sequence  
contigs within a contig-scaffold that consist entirely of whole  
genome shotgun sequence reads. Both end sequences and whole genome  
shotgun sequence only contigs will be indicated in the feature  
table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: http://www.hgsc.bcm.tmc.edu/  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GWUN  
Center clone name: CH230-3215  
----- Summary Statistics  
Assembly program: Phrap; version 0.990329  
Consensus quality: 191191 bases at least Q40  
Consensus quality: 194508 bases at least Q30  
Consensus quality: 195878 bases at least Q20  
Estimated insert size: 227256; sum-of-contigs estimation  
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length  
\* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 4 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 330403: contig of 330403 bp in length  
\* 330404 330503: gap of unknown length  
\* 330504 331603: contig of 1100 bp in length  
\* 331604 331703: gap of unknown length  
\* 331704 331738: contig of 6035 bp in length  
\* 331739 337838: gap of unknown length  
\* 337839 345098: contig of 7260 bp in length.

FEATURES  
source  
1.345098  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-3215"

misc\_feature complement(177532..178431)  
/note="Clone boundary  
clone\_end:Sp6  
site:  
end\_sequence:BH328475"  
179290..181054  
/note="wgs\_end\_extension  
clone\_end:Sp6"

## ORIGIN

Query Match 39.8%; Score 34.6; DB 2; Length 345098;  
Best Local Similarity 70.6%; Pred. No. 0.21;  
Matches 60; Conservative 0; Mismatches 24; Indels 1; Gaps 1;  
QY 4 GAACCATCAATTAATGAAGTAGAGATTCTTGCGCTGAGACTTGACGAGGAGCAAG-AAAG 62  
DB 149860 GAAGCAGCCTTGCTTGAGCAGAGCTCCCTTGAACCTTAAGACTTTGAGGAAACAAAGAAAG 149801  
QY 63 ACACTGAGACACCACTATGACAG 87  
DB 149800 TCAAGCTGAGAGCCGAATGACAG 149776

Search completed: March 12, 2004, 21:22:29  
Job time: 277.899 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 12, 2004, 12:26:39 ; Search time 34.8377 Seconds

(without alignments)  
10609.014 Million cell updates/sec

Title: US-09-966-880a-11

Perfect score: 87  
Sequence: 1 agagaccatcattatgta.....ctggacacacactatgacag 87

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 3373863 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : N Geneseq\_29Jan04:\*

1: geneseqn1980s:\*\n2: geneseqn1990s:\*\n3: geneseqn2000s:\*\n4: geneseqn2001as:\*\n5: geneseqn2001bs:\*\n6: geneseqn2002s:\*\n7: geneseqn2003as:\*\n8: geneseqn2003bs:\*\n9: geneseqn2003cs:\*\n10: geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	87	100.0	87	3	AAC55315 Human act
2	87	100.0	2818	3	AAC55312 Human act
3	87	100.0	5514	3	AAC55313 Human act
4	87	100.0	11204	3	AAC55339 Human act
5	87	100.0	11204	6	AB873286 DNA encod
6	84	96.6	1543	7	ABX05468 Human nov
7	84	96.6	2791	6	AB873287 DNA encod
8	84	96.6	2791	6	AB873288 DNA encod
9	41	47.1	2440	3	AAC55307 Mouse act
10	30.6	35.2	42299	4	AAK68932 Human imm
11	30	34.5	328	6	AB097746 Mouse ES
12	28.8	33.1	762	2	AAZ16628 Human gen
13	28.2	32.4	958	2	AAV33481 Human dis
14	27.8	32.0	2716	9	ADD48661 Rat gene
15	27.6	31.7	444	9	ADD32481 Human mit
16	27.6	31.7	496	4	AA185721 Human kin
17	27.6	31.7	515	8	AA857351 Human kin
18	27.6	31.7	843	2	AA731065 Human dec
19	27.6	31.5	1371	10	ADE28378 Human KFP
20	27.4	31.5	947	7	AA85023 Human sec
21	27.4	31.3	5810	5	AA84731 DNA encod
22	27.2	31.3	10497	4	AAK67302 Human imm
23	27.2	31.3	10497	4	AAK67302 Human imm

C	24	27.2	31.3	10497	4	AAK83166 Human imm
	25	27	31.0	389	2	AA186461 Human pol
	26	26.4	30.3	360	2	AAQ61230 Human bra
	27	26.2	30.1	9353	7	ABZ68800 Nucleotid
	28	26.2	30.1	10597	7	ABZ68799 Nucleotid
	29	26.2	30.1	10953	7	ABZ68805 Nucleotid
C	30	26.2	30.1	32154	4	AAJ37068 Human mus
	31	26.2	30.1	32154	4	ABK60056 CDNA enco
	32	26	29.9	26	3	AAC55335 Human act
	33	26	29.9	412	6	ABZ35592 Human gen
	34	26	29.9	417	7	ABZ71602 Breast sp
	35	26	29.9	444	7	ACF56780 Rice endo
	36	26	29.9	1328	7	ABZ71603 Breast sp
	37	26	29.9	2004	4	ABL10849 Drosophi1
C	38	26	29.9	2382	4	AAH18091 Human cdn
C	39	26	29.9	2386	9	ADD29848 Human tum
C	40	26	29.9	2772	4	AAK85945 Human imm
C	41	25.8	29.7	543	3	AAK08238 Human sec
C	42	25.8	29.7	543	6	ABK36224 CDNA sequ
	43	25.8	29.7	2506	4	ABL12873 Drosophi1
	44	25.8	29.7	4117	4	ABL12873 Drosophi1
	45	25.8	29.7	4699	4	ABL12872 Drosophi1

## ALIGNMENTS

RESULT 1	AAK55315 standard; DNA; 87 BP.
ID	AAK55315
AC	AAK55315;
XX	
DT	05-FEB-2001 (first entry)
XX	
DE	Human activation-induced cytidine deaminase exon 1 SEQ ID NO:11.
XX	
KW	Activation-induced cytidine deaminase; AID; cytidine deaminase;
KW	immune related disease; allergy; allergic disease; antiallergic;
KW	antianaemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;
KW	gene therapy; B cell associated immune system disorder; food allergy;
KW	immunodeficiency disease; immunoglobulin A deficiency disease; asthma;
KW	IGA nephritis; gamma-globulinemia; atopic dermatitis; allergic colitis;
KW	drug allergy; allergic rhinitis; Rosen disease; Digeorge disease; AIDS;
KW	ataxia telangiectasia; common variable immunodeficiency disorder;
KW	major histocompatibility class II deficiency disease;
KW	auto immunodeficiency syndrome; IgG subclass selection disorder; ds.
XX	
OS	Homo sapiens.
XX	
PN	W0200058480-A1.
XX	
PD	05-OCT-2000.
XX	
PF	28-MAR-2000; 2000MO-JP001918.
XX	
PR	29-MAR-1999; 99JP-00087192.
PR	24-JUN-1999; 99JP-00178999.
PR	27-DEC-1999; 99JP-00371382.
XX	
PA	(NIBS) JAPAN TOBACCO INC.
PA	(HONT/) HONTO T.
XX	
PI	Honjo T, Matsumatsu M;
XX	
DR	WPI; 2000-611715/58.
XX	
PT	Nucleic acid encoding activation induced cytidine deaminase, useful as a
PT	target for drug development for immune-related diseases including
PT	allergies.
XX	
PS	Claim 18; Page 150; 174pp; Japanese.
XX	



CC The present invention describes an activation-induced cytidine deaminase  
 CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and has  
 CC cytidine activity similar to APOBEC-1. AID has anti-allergic, anti-naemic,  
 CC antiasthmatic, ophthalmological, anti-HIV and dermatological activities,  
 CC and can be used in gene therapy. AID polynucleotides are useful in  
 CC methods for identifying drugs for the treatment of B cell associated  
 CC immune system disorders, immunodeficiency diseases and allergies, such as  
 CC immunoglobulin A (IgA) deficiency disease, IgA nephritis, gamma-  
 CC globulinemia, atopic dermatitis, allergic colitis, asthma, food allergy,  
 CC drug allergy, allergic rhinitis, Rosen disease, DisGeorge disease, ataxia  
 CC telangiectasia, common variable immunodeficiency disorder, MHC (major  
 CC histocompatibility class II deficiency disease, AIDS (auto  
 CC immunodeficiency syndrome), elevated Ige disorder, and Igg subclasses  
 CC selection disorder. The DNA sequences encoding AID may be used for gene  
 CC therapy and the antibodies to the AID protein may be used for diagnosis  
 CC and treatment of these disorders. The present sequence represents the  
 CC exon 1 genomic DNA sequence of human AID

CC Sequence 87 BP; 28 A; 17 C; 23 G; 19 T; 0 U; 0 Other;

CC Query Match 100.0%; Score 87; DB 3; Length 87;

CC Best Local Similarity 100.0%; Pred. No. 3.5e-22;

CC Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CC 1 AGAGAACCATCATTAATGAGAGATTTTCTGCGCTGAGACTTGAGAGGAGGACAGA 60  
 CC 1 AGAGAACCATCATTAATGAGAGATTTTCTGCGCTGAGACTTGAGAGGAGGACAGA 60  
 CC 1 AGAGAACCATCATTAATGAGAGATTTTCTGCGCTGAGACTTGAGAGGAGGACAGA 60

CC 61 AGACACTCTGAGACACCATATGACAG 87  
 CC 61 AGACACTCTGAGACACCATATGACAG 87

CC 05-FEB-2001 (first entry)

CC Human activation-induced cytidine deaminase encoding cDNA SEQ ID NO:7.

CC Activation-induced cytidine deaminase; AID; cytidine deaminase;  
 CC immune related disease; allergy; allergic disease; anti-allergic;  
 CC antinaemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;  
 CC gene therapy; B cell associated immune system disorder; food allergy;  
 CC immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
 CC IgA nephritis; gamma-globulinemia; atopic dermatitis; allergic colitis;  
 CC drug allergy; allergic rhinitis; Rosen disease; DisGeorge disease; AIDS;  
 CC ataxia telangiectasia; common variable immunodeficiency disorder;  
 CC major histocompatibility class II deficiency disease;  
 CC auto immunodeficiency syndrome; Igg subclass selection disorder; ss.

CC Homo sapiens.

CC Key Location/Qualifiers  
 CC CDS 80..676  
 CC /tag= a  
 CC /product= "activation-induced cytidine deaminase"

CC WO200058480-A1.

CC 05-OCT-2000.

CC 28-MAR-2000; 2000WO-JP001918.

CC 29-MAR-1999; 99JP-00087192.  
 CC 24-JUN-1999; 99JP-00178999.  
 CC 27-DEC-1999; 99JP-003771382.

CC (NIBS ) JAPAN TOBACCO INC.  
 CC (HONJ/) HONJO T.

XX Honjo T. Muramatsu M.  
 XX WPI, 2000-611715/58.  
 XX P-PDB; AAB24198.  
 XX Nucleic acid encoding activation induced cytidine deaminase, useful as a  
 XX target for drug development for immune-related diseases including  
 XX allergies.

XX Claim 3; Page 135-139; 174pp; Japanese.

XX The present sequence encodes human activation-induced cytidine deaminase  
 XX (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and has  
 XX cytidine activity similar to APOBEC-1. AID has anti-allergic, antinaemic,  
 XX antiasthmatic, ophthalmological, anti-HIV and dermatological activities,  
 XX and can be used in gene therapy. AID polynucleotides are useful in  
 XX methods for identifying drugs for the treatment of B cell associated  
 XX immune system disorders, immunodeficiency diseases and allergies, such as  
 XX immunoglobulin A (IgA) deficiency disease, IgA nephritis, gamma-  
 XX globulinemia, atopic dermatitis, allergic colitis, asthma, food allergy,  
 XX drug allergy, allergic rhinitis, Rosen disease, DisGeorge disease, ataxia  
 XX telangiectasia, common variable immunodeficiency disorder, MHC (major  
 XX histocompatibility class II deficiency disease, AIDS (auto  
 XX immunodeficiency syndrome), elevated Ige disorder, and Igg subclasses  
 XX selection disorder. The DNA sequences encoding AID may be used for gene  
 XX therapy and the antibodies to the AID protein may be used for diagnosis  
 XX and treatment of these disorders

XX Sequence 2818 BP; 868 A; 548 C; 626 G; 776 T; 0 U; 0 Other;

XX Query Match 100.0%; Score 87; DB 3; Length 2818;

XX Best Local Similarity 100.0%; Pred. No. 1.1e-21;

XX Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX 1 AGAGAACCATCATTAATGAGAGATTTTCTGCGCTGAGACTTGAGAGGAGGACAGA 60  
 XX 1 AGAGAACCATCATTAATGAGAGATTTTCTGCGCTGAGACTTGAGAGGAGGACAGA 60  
 XX 1 AGAGAACCATCATTAATGAGAGATTTTCTGCGCTGAGACTTGAGAGGAGGACAGA 60

XX 61 AGACACTCTGAGACACCATATGACAG 87  
 XX 61 AGACACTCTGAGACACCATATGACAG 87

XX 05-FEB-2001 (first entry)

XX Human activation-induced cytidine deaminase genomic DNA SEQ ID NO:9.

XX Activation-induced cytidine deaminase; AID; cytidine deaminase;  
 XX immune related disease; allergy; allergic disease; anti-allergic;  
 XX antinaemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;  
 XX gene therapy; B cell associated immune system disorder; food allergy;  
 XX immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
 XX IgA nephritis; gamma-globulinemia; atopic dermatitis; allergic colitis;  
 XX drug allergy; allergic rhinitis; Rosen disease; DisGeorge disease; AIDS;  
 XX ataxia telangiectasia; common variable immunodeficiency disorder;  
 XX major histocompatibility class II deficiency disease;  
 XX auto immunodeficiency syndrome; Igg subclass selection disorder; ds.

XX Homo sapiens.

XX WO200058480-A1.

XX 05-OCT-2000.

XX 28-MAR-2000; 2000WO-JP001918.



PR 29-MAR-1999; 99JP-00087192.  
 PR 24-JUN-1999; 99JP-00178999.  
 PR 27-DEC-1999; 99JP-00371382.  
 XX  
 PA (NIBS) JAPAN TOBACCO INC.  
 PA (HONJ/) HONJO T.  
 PI Honjo T, Muramatsu M;  
 DR WPI; 2000-611715/58.  
 XX  
 PT Nucleic acid encoding activation induced cytidine deaminase, useful as a  
 PT target for drug development for immune-related diseases including  
 PT allergies.  
 XX  
 PS Claim 17; Page 142-145; 174pp; Japanese.  
 XX  
 CC The present invention describes an activation-induced cytidine deaminase  
 CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and has  
 CC cytidine activity similar to APOBEC-1. AID has antiallergic, antianemic,  
 CC antiasthmatic, ophthalmological, anti-HIV and dermatological activities,  
 CC and can be used in gene therapy. AID polynucleotides are useful in  
 CC methods for identifying drugs for the treatment of B cell associated  
 CC immune system disorders, immunodeficiency diseases and allergies, such as  
 CC immunoglobulin A (IGA) deficiency disease, IGA nephritis, gamma-  
 CC globulinemia, atopic dermatitis, allergic colitis, asthma, food allergy,  
 CC drug allergy, allergic rhinitis, Rosen disease, Digorge disease, ataxia  
 CC telangiectasia, common variable immunodeficiency disorder, MHC (major  
 CC histocompatibility class II deficiency disease, AIDS (auto  
 CC immunodeficiency syndrome), elevated IGE disorder, and IGG subclass  
 CC selection disorder. The DNA sequences encoding AID may be used for gene  
 CC therapy and the antibodies to the AID protein may be used for diagnosis  
 CC and treatment of these disorders. The present sequence represents a  
 CC genomic DNA sequence of human AID  
 CC  
 XX  
 SQ Sequence 5514 BP; 1709 A; 1045 C; 1134 G; 1623 T; 0 U; 3 Other;  
 Query Match 100.0%; Score 87; DB 3; Length 5514;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-21;  
 Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AGAGAACCATCATTTAATTGAGTGAAGTATTTCTGGCCTGAGACTTGACGGAGGCAAGA 60  
 DB 1032 AGAGAACCATCATTTAATTGAGTGAAGTATTTCTGGCCTGAGACTTGACGGAGGCAAGA 1091  
 QY 61 AGACACTCTGGACACCACTATGACAG 87  
 DB 1092 AGACACTCTGGACACCACTATGACAG 1118  
 RESULT 4  
 AAC55339  
 ID AAC55339 standard; DNA; 11204 BP.  
 XX  
 AC AAC55339;  
 XX  
 DT 05-FEB-2001 (first entry)  
 XX  
 DE Human activation-induced cytidine deaminase genomic DNA SEQ ID NO:35.  
 XX  
 XX Activation-induced cytidine deaminase; AID; cytidine deaminase;  
 KW immune related disease; allergy; allergic disease; antiallergic;  
 KW antianemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;  
 KW gene therapy; B cell associated immune system disorder; food allergy;  
 KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
 KW IGA nephritis; gamma-globulinemia; atopic dermatitis; allergic colitis;  
 KW drug allergy; allergic rhinitis; Rosen disease; Digorge disease; AIDS;  
 KW ataxia telangiectasia; common variable immunodeficiency disorder;  
 KW major histocompatibility class II deficiency disease;  
 KW auto immunodeficiency syndrome; IGG subclass selection disorder; ds.  
 XX  
 OS Homo sapiens.  
 XX

PN WO200058480-A1.  
 XX  
 PD 05-OCT-2000.  
 XX  
 XX 28-MAR-2000; 2000WO-JP001918.  
 XX  
 XX 29-MAR-1999; 99JP-00087192.  
 PR 24-JUN-1999; 99JP-00178999.  
 PR 27-DEC-1999; 99JP-00371382.  
 XX  
 PA (NIBS) JAPAN TOBACCO INC.  
 PA (HONJ/) HONJO T.  
 PI Honjo T, Muramatsu M;  
 DR WPI; 2000-611715/58.  
 XX  
 XX Nucleic acid encoding activation induced cytidine deaminase, useful as a  
 PT target for drug development for immune-related diseases including  
 PT allergies.  
 XX  
 PS Claim 17; Page 163-170; 174pp; Japanese.  
 XX  
 CC The present invention describes an activation-induced cytidine deaminase  
 CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and has  
 CC cytidine activity similar to APOBEC-1. AID has antiallergic, antianemic,  
 CC antiasthmatic, ophthalmological, anti-HIV and dermatological activities,  
 CC and can be used in gene therapy. AID polynucleotides are useful in  
 CC methods for identifying drugs for the treatment of B cell associated  
 CC immune system disorders, immunodeficiency diseases and allergies, such as  
 CC immunoglobulin A (IGA) deficiency disease, IGA nephritis, gamma-  
 CC globulinemia, atopic dermatitis, allergic colitis, asthma, food allergy,  
 CC drug allergy, allergic rhinitis, Rosen disease, Digorge disease, ataxia  
 CC telangiectasia, common variable immunodeficiency disorder, MHC (major  
 CC histocompatibility class II deficiency disease, AIDS (auto  
 CC immunodeficiency syndrome), elevated IGE disorder, and IGG subclass  
 CC selection disorder. The DNA sequences encoding AID may be used for gene  
 CC therapy and the antibodies to the AID protein may be used for diagnosis  
 CC and treatment of these disorders. The present sequence represents a  
 CC genomic DNA sequence of human AID  
 CC  
 XX  
 SQ Sequence 11204 BP; 3305 A; 2273 C; 2373 G; 3253 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 87; DB 3; Length 11204;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-21;  
 Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AGAGAACCATCATTTAATTGAGTGAAGTATTTCTGGCCTGAGACTTGACGGAGGCAAGA 60  
 DB 442 AGAGAACCATCATTTAATTGAGTGAAGTATTTCTGGCCTGAGACTTGACGGAGGCAAGA 501  
 QY 61 AGACACTCTGGACACCACTATGACAG 87  
 DB 502 AGACACTCTGGACACCACTATGACAG 528  
 RESULT 5  
 ABS73286  
 ID ABS73286 standard; DNA; 11204 BP.  
 XX  
 AC ABS73286;  
 XX  
 DT 04-DEC-2002 (first entry)  
 XX  
 DE DNA encoding human translocation del(12p) protein #1.  
 XX  
 KW Chromosome aberration; oncogenic fusion protein; cancer;  
 KW proliferative disease; cellular protein isoform; heat shock protein 90;  
 KW HSP-90; rheumatoid arthritis; cancer; haematopoietic disorder;  
 KW T cell lymphoma; B cell lymphoma; chronic myeloid leukaemia; CML;  
 KW acute myeloid leukaemia; ALL; chronic myelomonocytic leukaemia; CMML;  
 KW acute lymphoblastic leukaemia; ALL; APL; NHL; solid tumour;  
 KW papillary thyroid carcinoma; Ewing's sarcoma; melanoma; liposarcoma;  
 KW



XX	thadomyosarcoma; synovial sarcoma; viral infection; gene; ds.	
OS	Homo sapiens.	
XX	WO200269900-A2.	
XX	12-SEP-2002.	
PD		
XX		
PF	01-MAR-2002; 2002WO-US006518.	
XX		
PR	01-MAR-2001; 2001US-0272751P.	
XX	(CONF-) CONFORMA THERAPEUTICS CORP.	
PA		
XX	Fritz LC, Burrows FU;	
XX	WPI; 2002-698710/75.	
XX	P-PSDB; ABG95082.	
DR		
XX	Treating genetically-defined disease associated with chromosomal	
PT	aberrations yielding oncogenic fusion proteins; e.g. cell proliferative	
XX	diseases, involves administering an inhibitor of heat shock protein 90.	
XX		
PS	Disclosure; Page 242-245; 389pp: English.	
XX		
XX	The invention describes a method of treating genetically-defined disease	
CC	associated with chromosomal aberrations yielding oncogenic fusion	
CC	proteins (I), treating cancerous cells containing (I) in a heterogeneous	
CC	cell population, treating proliferative diseases associated with mutant	
CC	protein or cellular protein isoforms (II) dependent on heat shock protein	
CC	(HSP)-90, or selectively treating cells expressing (II) involving	
CC	administering HSP90-inhibitor. The method is useful for treating	
CC	genetically-defined disease with chromosomal aberration yielding	
CC	oncogenic fusion protein, treating cancerous cells containing fusion	
CC	protein in heterogeneous cell population, treating proliferative disease	
CC	(e.g. rheumatoid arthritis or cancer) associated with mutant protein or	
CC	cellular protein isoform dependent on heat shock protein (HSP)-90 (e.g.	
CC	p53), or selectively treating cells expressing mutant protein or cellular	
CC	protein isoform in a patient heterozygous for (II). The method is useful	
CC	for treating a disease e.g. haematopoietic disorder such as T or B cell	
CC	lymphoma, chronic myeloid leukaemia (CML), APL, ALL, AML, NHL and CMM, or	
CC	a disease characterised by a solid tumour such as papillary thyroid	
CC	carcinoma, Ewing's sarcoma, melanoma, liposarcoma, rhabdomyosarcoma and	
CC	synovial sarcoma. The method is also useful for treating viral	
CC	infections. This represents the DNA sequence of a chromosome aberration	
XX		
SQ	Sequence 11204 BP; 3305 A; 2273 C; 2373 G; 3253 T; 0 U; 0 Other;	
XX		
QY	Query Match 100.0%; Score 87; DB 6; Length 11204;	
DB	Best Local Similarity 100.0%; Pred. No. 1.8e-21;	
XX	Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0	
QY	1 AGAGAACCTATTATTTAGTGAAGATTTTTCTGGCCCGAGACTTGGAGGAGGAGAGA 60	
DB	442 AGAGAACCTATTATTTAGTGAAGATTTTTCTGGCCCGAGACTTGGAGGAGGAGAGA 501	
QY	61 AGACACTGTGAGACCACTATTGGACAG 87	
DB	502 AGACACTGTGAGACCACTATTGGACAG 528	

RESULT 6  
ABX05468  
ID ABX05468 strand: cDNA, 1543 BP.  
XX  
AC ABX05468;  
XX  
XX 17-JAN-2003 (first entry)  
DT  
XX  
DE Human novel polynucleotide #483.  
XX  
KW Human; gene; ss; genetic disorder; gene mapping; medical imaging; cancer  
KW neurodegenerative disorder; lymphoid cell disorder; osteoporosis;

OS Parkinson's disease; Alzheimer's disease; bone degenerative disorder;  
XX osteoarthritis; periodontal disease; liver fibrosis; viral infection;  
XX fungal infection; bacterial infection; autoimmune disease; diabetes;  
XX atopic dermatitis.  
XX Homo sapiens.  
XX  
XX WO200274961-A1.  
PN  
XX 26-SEP-2002.  
PD  
XX  
PF 14-MAR-2002; 2002WO-US005109.  
PR 15-MAR-2001; 2001US-00810173.  
XX  
PA (HXS-) HXSEQ INC.  
XX  
PI Tang YT, Zhou P, Goodrich R, Asundi V, Zhang J, Zhao Q, Ren F;  
PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M,  
PI Wehrman T, Wang J, Wang D, Drmanac RT,  
XX WPI; 2003-040556/03.  
DR P-PSDB; ABU00390.  
XX  
XX New isolated polypeptides and polynucleotides, useful for preventing,  
PT treating or ameliorating medical conditions, such as cancer,  
PT neurodegenerative disorders, lymphoid cell disorders, bone degenerative  
PT disorders, and infections.  
XX  
XX Claim 1; SEQ ID NO 483; 235pp; English.

The invention relates to human polynucleotides and the polypeptides they encode. The polynucleotides and polypeptides are useful in diagnostics, forensics, gene mapping, medical imaging, identification of mutations responsible for genetic disorders or other traits, assessing biodiversity and producing many other types of data and products dependent on DNA and amino acid sequences. They are also useful for preventing, treating or ameliorating medical conditions, such as cancer, neurodegenerative disorders (e.g. Parkinson's disease, Alzheimer's disease), lymphoid cell disorders, osteoporosis, osteoarthritis, bone degenerative disorders, periodontal disease, liver fibrosis, infections (e.g. viral, fungal or bacterial) or autoimmune diseases (e.g. diabetes, atopic dermatitis). Sequences ABX04966-ABX0511 represent human polynucleotides of the invention. Note: The sequence data for this patent is not represented in the printed specification but is based on sequence information supplied by the European Patent Office

Sequence 1543 BP, 428 A, 334 C, 352 G, 429 T, 0 U, 0 Other;  
SQ

Query Match 96.6%; Score 84; DB 7; Length 1543;  
Best Local Similarity 100.0%; Pred. No. 1.2e-20;  
Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GAACATCATTAATTGAGTGAAGATTTTCGGCCTGAGACTGTGAGGAGCAAGAAGA 63  
| | | | |  
DB 2 GAACATCATTAATTGAGTGAAGATTTTCGGCCTGAGACTGTGAGGAGCAAGAAGA 61  
| | | | |  
QY 64 CACTCTGCACACCACCATATGACAG 87  
| | | | |  
DB 62 CACTCTGCACACCACCATATGACAG 85  
| | | | |

RESULT 7	
AB573287	
ID	AB573287 standard; DNA; 2791 BP.
XX	
AC	AB573287;
XX	
DT	04-DEC-2002 (first entry)
XX	
DE	DNA encoding human translocation del(12p) protein #2.
XX	
KN	Chromosome aberration; oncogenic fusion protein; cancer



KM proliferative disease; cellular protein isoform; heat shock protein 90;  
 KM HSP-90; rheumatoid arthritis; cancer; haematopoietic disorder;  
 KM T cell lymphoma; B cell lymphoma; chronic myeloid leukaemia; CML;  
 KM acute myeloid leukaemia; AML; chronic myelomonocytic leukaemia; CMML;  
 KM acute lymphoblastic leukaemia; ALL; APL; NHL; solid tumour;  
 KM papillary thyroid carcinoma; Ewing's sarcoma; liposarcoma;  
 KM rhabdomyosarcoma; synovial sarcoma; viral infection; gene; ds.  
 XX  
 OS Homo sapiens.  
 PN WO200269900-A2.  
 XX  
 PD 12-SEP-2002.  
 XX  
 PF 01-MAR-2002; 2002WO-US06518.  
 XX  
 PR 01-MAR-2001; 2001US-0272751P.  
 XX  
 PA (CONF-) CONFORMA THERAPEUTICS CORP.  
 XX  
 PI Fritz LC, Burrows FU;  
 XX  
 DR WPI; 2002-698710/75.  
 DR P-PSDB; ABG95083.  
 XX  
 PT Treating genetically-defined disease associated with chromosomal  
 PT aberrations yielding oncogenic fusion proteins, e.g. cell proliferative  
 PT diseases, involves administering an inhibitor of heat shock protein 90.  
 XX  
 PS Disclosure; Page 246-247; 389pp; English.  
 XX  
 CC The invention describes a method of treating genetically-defined disease  
 CC associated with chromosomal aberrations yielding oncogenic fusion  
 CC proteins (I), treating cancerous cells containing (I) in a heterogeneous  
 CC cell population, treating proliferative diseases associated with mutant  
 CC protein or cellular protein isoforms (II) dependent on heat shock protein  
 CC (HSP)-90, or selectively treating cells expressing (II) involving  
 CC administering HSP90-inhibitor. The method is useful for treating  
 CC genetically-defined disease with chromosomal aberration yielding  
 CC oncogenic fusion protein, treating cancerous cells containing fusion  
 CC protein in heterogeneous cell population, treating proliferative disease  
 CC (e.g. rheumatoid arthritis or cancer) associated with mutant protein or  
 CC cellular protein isoform dependent on heat shock protein (HSP)-90 (e.g.  
 CC p53), or selectively treating cells expressing mutant protein or cellular  
 CC protein isoform in a patient heterozygous for (II). The method is useful  
 CC for treating a disease e.g. haematopoietic disorder such as T or B cell  
 CC lymphoma, chronic myeloid leukaemia (CML), APL, ALL, AML, NHL and CMML,  
 CC or a disease characterised by a solid tumour such as papillary thyroid  
 CC carcinoma, Ewing's sarcoma, melanoma, liposarcoma, rhabdomyosarcoma and  
 CC synovial sarcoma. The method is also useful for treating viral  
 CC infections. This represents the DNA sequence of a chromosome aberration  
 XX  
 SQ Sequence 2791 BP; 842 A; 548 C; 625 G; 776 T; 0 U; 0 Other;  
 Query Match 96.6%; Score 84; DB 6; Length 2791;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-20;  
 Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 4 GAACCATCATTAATTGAAGTGAATTTTCTGGCCTGAGACTTGCGAGGAGCAAGAAA 63  
 DB 1 GAACCATCATTAATTGAAGTGAATTTTCTGGCCTGAGACTTGCGAGGAGCAAGAAA 60  
 QY 64 CACTCTGGACACCACTATGACAG 87  
 DB 61 CACTCTGGACACCACTATGACAG 84

RESULT 8  
 AB873288  
 ID AB873288 standard; DNA; 2791 BP.  
 AC  
 XX AB873288;  
 XX

DT 04-DEC-2002 (first entry)  
 XX  
 DE DNA encoding human translocation del(12p) protein #3.  
 XX  
 KM Chromosome aberration; oncogenic fusion protein; cancer;  
 KM proliferative disease; cellular protein isoform; heat shock protein 90;  
 KM HSP-90; rheumatoid arthritis; cancer; haematopoietic disorder;  
 KM T cell lymphoma; B cell lymphoma; chronic myeloid leukaemia; CML;  
 KM acute myeloid leukaemia; AML; chronic myelomonocytic leukaemia; CMML;  
 KM acute lymphoblastic leukaemia; ALL; APL; NHL; solid tumour;  
 KM papillary thyroid carcinoma; Ewing's sarcoma; melanoma; liposarcoma;  
 KM rhabdomyosarcoma; synovial sarcoma; viral infection; gene; ds.  
 XX  
 OS Homo sapiens.  
 PN WO200269900-A2.  
 XX  
 PD 12-SEP-2002.  
 XX  
 PF 01-MAR-2002; 2002WO-US06518.  
 XX  
 PR 01-MAR-2001; 2001US-0272751P.  
 XX  
 PA (CONF-) CONFORMA THERAPEUTICS CORP.  
 XX  
 PI Fritz LC, Burrows FU;  
 XX  
 DR WPI; 2002-698710/75.  
 DR P-PSDB; ABG95084.  
 XX  
 PT Treating genetically-defined disease associated with chromosomal  
 PT aberrations yielding oncogenic fusion proteins, e.g. cell proliferative  
 PT diseases, involves administering an inhibitor of heat shock protein 90.  
 XX  
 PS Disclosure; Page 248-249; 389pp; English.  
 XX  
 CC The invention describes a method of treating genetically-defined disease  
 CC associated with chromosomal aberrations yielding oncogenic fusion  
 CC proteins (I), treating cancerous cells containing (I) in a heterogeneous  
 CC cell population, treating proliferative diseases associated with mutant  
 CC protein or cellular protein isoforms (II) dependent on heat shock protein  
 CC (HSP)-90, or selectively treating cells expressing (II) involving  
 CC administering HSP90-inhibitor. The method is useful for treating  
 CC genetically-defined disease with chromosomal aberration yielding  
 CC oncogenic fusion protein, treating cancerous cells containing fusion  
 CC protein in heterogeneous cell population, treating proliferative disease  
 CC (e.g. rheumatoid arthritis or cancer) associated with mutant protein or  
 CC cellular protein isoform dependent on heat shock protein (HSP)-90 (e.g.  
 CC p53), or selectively treating cells expressing mutant protein or cellular  
 CC protein isoform in a patient heterozygous for (II). The method is useful  
 CC for treating a disease e.g. haematopoietic disorder such as T or B cell  
 CC lymphoma, chronic myeloid leukaemia (CML), APL, ALL, AML, NHL and CMML,  
 CC or a disease characterised by a solid tumour such as papillary thyroid  
 CC carcinoma, Ewing's sarcoma, melanoma, liposarcoma, rhabdomyosarcoma and  
 CC synovial sarcoma. The method is also useful for treating viral  
 CC infections. This represents the DNA sequence of a chromosome aberration  
 XX  
 SQ Sequence 2791 BP; 842 A; 548 C; 625 G; 776 T; 0 U; 0 Other;  
 Query Match 96.6%; Score 84; DB 6; Length 2791;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-20;  
 Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 4 GAACCATCATTAATTGAAGTGAATTTTCTGGCCTGAGACTTGCGAGGAGCAAGAAA 63  
 DB 1 GAACCATCATTAATTGAAGTGAATTTTCTGGCCTGAGACTTGCGAGGAGCAAGAAA 60  
 QY 64 CACTCTGGACACCACTATGACAG 87  
 DB 61 CACTCTGGACACCACTATGACAG 84

RESULT 9



ID	AAC55307
XX	AAC55307 standard; cDNA; 2440 BP.
AC	AAC55307;
XX	
DT	05-FEB-2001 (first entry)
DE	Mouse activation-induced cytidine deaminase encoding cDNA SEQ ID NO:1.
XX	
KW	Activation-induced cytidine deaminase; AID; cytidine deaminase;
KW	immune related disease; allergy; allergic disease; antiallergic;
KW	antianemic; antihistaminic; ophthalmological; anti-HIV; dermatologic;
KW	gene therapy; B cell associated immune system disorder; food allergy;
KW	immunodeficiency disease; immunoglobulin A deficiency disease; asthma;
KW	Iga nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;
KW	drug allergy; allergic rhinitis; Rosen disease; DiGeorge disease; AIDS;
KW	ataxia telangiectasia; common variable immunodeficiency disorder;
KW	major histocompatibility class II deficiency disease;
KW	auto immunodeficiency syndrome; Igs subclass selection disorder; ss.
XX	
OS	Mus musculus.
XX	
FH	Key Location/Qualifiers
FT	CDS 93..689
FT	/tag= a
PT	/product= "activation-induced cytidine deaminase"
XX	
PN	WO200058480-A1.
XX	
PD	05-OCT-2000.
XX	
PF	28-MAR-2000; 2000MO-JP001918.
XX	
PR	29-MAR-1999; 99GP-00087192.
PR	24-JUN-1999; 99GP-00178999.
PR	27-DEC-1999; 99GP-00371382.
XX	
PA	(NIBS) JAPAN TOBACCO INC.
PA	(HONBU) HONDO T.
PI	Honjo T, Muramatsu M;
XX	
DR	WPI; 2000-611715/58.
DR	P-PDB; AAB24157.
XX	
PT	Nucleic acid encoding activation induced cytidine deaminase, useful as a target for drug development for immune-related diseases including allergies.
XX	
PS	Claim 3; Page 126-130; 174pp; Japanese.
XX	
CC	The present sequence encodes mouse activation-induced cytidine deaminase (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and has cytidine activity similar to APOBEC-1. AID has anti-allergic, antianemic, antihistaminic, ophthalmological, anti-HIV and dermatological activities, and can be used in gene therapy. AID polynucleotides are useful in methods for identifying drugs for the treatment of B cell associated immune system disorders, immunodeficiency diseases and allergies, such as immunoglobulin A (Iga) deficiency disease, Iga nephritis, gamma-globulinaemia, atopic dermatitis, allergic colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen disease, DiGeorge disease, ataxia telangiectasia, common variable immunodeficiency disorder, MHC (major histocompatibility class II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated Igs disorder, and Igs subclasses selection disorder. The DNA sequences encoding AID may be used for gene therapy and the antibodies to the AID protein may be used for diagnosis and treatment of these disorders
XX	
SQ	Sequence 2440 BP; 706 A; 546 C; 551 G; 636 T; 0 U; 1 Other;
XX	
Query Match	47.1%; Score 41; DB 3; Length 2440;
Best Local Similarity	75.3%; Pred. No. 0.00011;
Matches 64; Conservative	0; Mismatches 20; Indels 1; Gaps 1

Cy		GAGCCATCATTAATTGAAGAATTTTCTGGCGTAGACTTG-CAGGAGGCAG-AAG	62
D6		16 GAGCACCCTTGCTGCTGAGCAGACCTTCCTTGCCCTAAGACTTTGAGGAGTCAGAAAG	75
Cy	63	ACACTGTGACACCACTATGTGACAG	87
D6	76	TCAAGCTGAGACCGATATGACAG	100
RESULT 10			
ID	AAK68932		
ID	AAK68932	standard; DNA; 42299 BP.	
XX	AAK68932;		
AC			
XX			
DT	06-NOV-2001	(first entry)	
XX			
DE	Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23744.		
XX			
KW	Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;		
KM	cyclostatic; gene therapy; vaccine; metastasis; ds.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200157182-A2.		
XX			
PD	09-AUG-2001.		
XX			
Pf	17-JAN-2001; 2001KO-US001354.		
XX			
PR	31-JAN-2000; 2000US-0179065P.		
PR	04-FEB-2000; 2000US-0180628P.		
PR	24-FEB-2000; 2000US-0184664P.		
PR	02-MAR-2000; 2000US-0186350P.		
PR	16-MAR-2000; 2000US-0189874P.		
PR	17-MAR-2000; 2000US-019076P.		
PR	18-APR-2000; 2000US-0198123P.		
PR	19-MAY-2000; 2000US-020515P.		
PR	07-JUN-2000; 2000US-0209467P.		
PR	28-JUN-2000; 2000US-0214886P.		
PR	30-JUN-2000; 2000US-0215135P.		
PR	07-JUL-2000; 2000US-0216647P.		
PR	11-JUL-2000; 2000US-0216880P.		
PR	07-JUL-2000; 2000US-0217487P.		
PR	11-JUL-2000; 2000US-0217496P.		
PR	14-JUL-2000; 2000US-0218290P.		
PR	26-JUL-2000; 2000US-0220963P.		
PR	26-JUL-2000; 2000US-0220964P.		
PR	14-AUG-2000; 2000US-0224518P.		
PR	14-AUG-2000; 2000US-0224519P.		
PR	14-AUG-2000; 2000US-0225213P.		
PR	14-AUG-2000; 2000US-0225214P.		
PR	14-AUG-2000; 2000US-0225266P.		
PR	14-AUG-2000; 2000US-0225267P.		
PR	14-AUG-2000; 2000US-0225268P.		
PR	14-AUG-2000; 2000US-0225270P.		
PR	14-AUG-2000; 2000US-0225447P.		
PR	14-AUG-2000; 2000US-0225757P.		
PR	14-AUG-2000; 2000US-0225758P.		
PR	14-AUG-2000; 2000US-0225759P.		
PR	18-AUG-2000; 2000US-0226279P.		
PR	22-AUG-2000; 2000US-0226681P.		
PR	22-AUG-2000; 2000US-0226868P.		
PR	22-AUG-2000; 2000US-0227182P.		
PR	23-AUG-2000; 2000US-0227009P.		
PR	30-AUG-2000; 2000US-0228924P.		
PR	01-SEP-2000; 2000US-0229287P.		
PR	01-SEP-2000; 2000US-0229343P.		
PR	01-SEP-2000; 2000US-0229344P.		
PR	01-SEP-2000; 2000US-0229345P.		
PR	05-SEP-2000; 2000US-0229509P.		
PR	05-SEP-2000; 2000US-0229513P.		







XX Mouse ES cell related CDNA SEQ ID NO 1014.  
 DE AAZ16628/c  
 XX ID AAZ16628 standard; CDNA; 762 BP.  
 KM Mouse; ES cell; gene trapped sequence; GTS; gene expression;  
 KM development disorder; cell differentiation disorder; gene; ss.  
 XX  
 OS Mus sp.  
 XX  
 PN US2002081668-A1.  
 XX  
 PD 27-JUN-2002.  
 XX  
 PF 30-NOV-2000; 2000US-00728446.  
 XX  
 PR 20-NOV-1998; 98US-0109302P.  
 PR 01-DEC-1999; 99US-0168270P.  
 XX  
 PA (FRIE/) FRIEDRICH G.  
 PA (ZAMB/) ZAMBROWICZ B.  
 PA (SAND/) SANDS A T.  
 XX  
 PI Friedrich G, Zambrowicz B, Sands AT;  
 XX  
 DR WPI; 2002-626541/67.  
 XX  
 PT Novel murine polynucleotides that individually identify novel genes into  
 PT which a retroviral gene trap vector has been integrated, useful in  
 PT genomic analysis and in discovery, development of therapeutic and  
 PT diagnostic agents.  
 XX  
 PS Claim 2; SEQ ID NO 1014; 23bp + Sequence Listing; English.  
 XX  
 CC The invention relates to isolated murine polynucleotides (I) comprising a  
 CC contiguous stretch of at least about 60 nucleotides of a sequence  
 CC (ABQ96733-ABQ98191) chosen from 1461 OMNIBANK gene trapped sequences  
 CC (GTSs). The novel genes can be used in a process to identify novel  
 CC polynucleotide sequences by comparing them to the novel gene sequences.  
 CC The novel genes and cells are useful in functional genomic analysis and  
 CC in the discovery and development of new therapeutic and diagnostic agents  
 CC and methods. (I) is useful for identifying the coding regions of the  
 CC murine genome, to isolate cDNAs, genomic clones or full-length  
 CC genes/polynucleotides or homologues, heterologues, paralogues or  
 CC orthologues that are capable of hybridizing to one or more of the GTSs  
 CC under stringent conditions. (I) can be incorporated into a phage display  
 CC system that can be used to screen for proteins or other ligands that are  
 CC capable of binding an amino acid sequence encoded by an oligonucleotide  
 CC or polynucleotide sequence in at least one of the GTS sequences. (I) is  
 CC useful in arrays, such as gene chips, to identify and characterize  
 CC temporal and tissue specific gene expression, to identify the gene of  
 CC interest from many sources and for genetic manipulations such as  
 CC antisense inhibition and gene targeting. Decreasing the level of  
 CC expression of (I) and/or down regulating the activity of peptides or  
 CC proteins encoded by (I) is useful for treating development and cell  
 CC differentiation disorders. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from USPTO at  
 CC seqdata.uspto.gov/sequence.html?docID=20020081668  
 XX  
 SQ Sequence 328 BP; 97 A; 74 C; 99 G; 56 T; 0 U; 2 Other;  
 XX  
 QY Query Match 34.5%; Score 30; DB 6; Length 328;  
 Db Best Local Similarity 59.3%; Pred. No. 0.63;  
 Matches 51; Conservative 0; Mismatches 35; Indels 0; Gaps 0;  
 QY 1 AGAGAACCATCATTTATGAGAGATTCTTGCGCTGAGACTTGACGAGGAGCAAGA 60  
 Db 103 AGAGAACCATCATTTATGAGAGATTCTTGCGCTGAGACTTGACGAGGAGCAAGA 60  
 QY 61 AGACACTTGGACACCCACTATGACA 86  
 Db 163 AACCGACCAAGTAGAATTCGGGAA 188

RESULT 12  
 XX AAZ16628/c  
 XX ID AAZ16628 standard; CDNA; 762 BP.  
 KM Mouse; ES cell; gene trapped sequence; GTS; gene expression;  
 KM development disorder; cell differentiation disorder; gene; ss.  
 XX  
 OS Mus sp.  
 XX  
 PN US2002081668-A1.  
 XX  
 PD 27-JUN-2002.  
 XX  
 PF 30-NOV-2000; 2000US-00728446.  
 XX  
 PR 20-NOV-1998; 98US-0109302P.  
 PR 01-DEC-1999; 99US-0168270P.  
 XX  
 PA (FRIE/) FRIEDRICH G.  
 PA (ZAMB/) ZAMBROWICZ B.  
 PA (SAND/) SANDS A T.  
 XX  
 PI Friedrich G, Zambrowicz B, Sands AT;  
 XX  
 DR WPI; 2002-626541/67.  
 XX  
 PT Novel murine polynucleotides that individually identify novel genes into  
 PT which a retroviral gene trap vector has been integrated, useful in  
 PT genomic analysis and in discovery, development of therapeutic and  
 PT diagnostic agents.  
 XX  
 PS Claim 2; SEQ ID NO 1014; 23bp + Sequence Listing; English.  
 XX  
 CC The invention relates to isolated murine polynucleotides (I) comprising a  
 CC contiguous stretch of at least about 60 nucleotides of a sequence  
 CC (ABQ96733-ABQ98191) chosen from 1461 OMNIBANK gene trapped sequences  
 CC (GTSs). The novel genes can be used in a process to identify novel  
 CC polynucleotide sequences by comparing them to the novel gene sequences.  
 CC The novel genes and cells are useful in functional genomic analysis and  
 CC in the discovery and development of new therapeutic and diagnostic agents  
 CC and methods. (I) is useful for identifying the coding regions of the  
 CC murine genome, to isolate cDNAs, genomic clones or full-length  
 CC genes/polynucleotides or homologues, heterologues, paralogues or  
 CC orthologues that are capable of hybridizing to one or more of the GTSs  
 CC under stringent conditions. (I) can be incorporated into a phage display  
 CC system that can be used to screen for proteins or other ligands that are  
 CC capable of binding an amino acid sequence encoded by an oligonucleotide  
 CC or polynucleotide sequence in at least one of the GTS sequences. (I) is  
 CC useful in arrays, such as gene chips, to identify and characterize  
 CC temporal and tissue specific gene expression, to identify the gene of  
 CC interest from many sources and for genetic manipulations such as  
 CC antisense inhibition and gene targeting. Decreasing the level of  
 CC expression of (I) and/or down regulating the activity of peptides or  
 CC proteins encoded by (I) is useful for treating development and cell  
 CC differentiation disorders. Note: The sequence data for this patent did  
 CC not form part of the printed specification, but was obtained in  
 CC electronic format directly from USPTO at  
 CC seqdata.uspto.gov/sequence.html?docID=20020081668  
 XX  
 SQ Sequence 762 BP; 212 A; 147 C; 163 G; 213 T; 0 U; 27 Other;  
 XX  
 QY Query Match 33.1%; Score 28.8; DB 2; Length 762;  
 Db Best Local Similarity 60.8%; Pred. No. 2.3;  
 Matches 45; Conservative 0; Mismatches 29; Indels 0; Gaps 0;



```

QY      4  GAACCATCATTAATGAAGTGAATTTTCTGGCCTGAGACTTGAGGGAGGCAAGA 63
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      409  GAAAGGTGGTTGGNAGACACNACATTCTTCTAGCTTCAGAAAGTCAGGAGCCCATCA 350
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      64  CACTCTGACACCA 77
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      349  CAGCCTGAACATCA 336
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
AAV33481/C
ID  AAV33481 standard; DNA; 958 BP.
XX
XX  AAV33481;
XX
XX  29-DEC-1998 (first entry)
XX
DB  Human disease related nucleotide kinase-1 (DRNK-1) DNA sequence.
XX
XX  Human disease related nucleotide kinase-1; DRNK-1; deoxyguanosine kinase;
KM  p21ras; cell proliferation; oncogenesis; cancer; gene therapy;
KM  immune disorder; neurological dysfunction; ss.
XX
XX  Homo sapiens.
XX
XX  Key Location/Qualifiers
FH  CDS 61..846
FT  /tag= a
FT  /product= "Human DRNK-1"
FT  /transit_peptide 61..135
FT  /tag= b
XX  /note= "Mitochondrial localisation signal"
XX
XX  US5817482-A.
XX
XX  06-OCT-1998.
XX
XX  20-JUN-1997; 97US-00879561.
XX
XX  20-JUN-1997; 97US-00879561.
XX
XX  (INCY-) INCYTE PHARM INC.
XX
XX  Hillman JL, Corley NC, Hawkins PR, Guegler KJ, Bandman O;
PI  WPI; 1998-556388/47.
DR  P-PSDB; AAW70493.
XX
XX  Nucleic acids encoding de:oxy:guanosine kinase - useful for recombinant
PT  production of the enzyme for treating diseases caused by lack of the
PT  enzyme e.g. cancers caused through loss of enzyme function.
XX
XX  Claim 4; Fig 1A-1C; 53pp; English.
XX
XX  The present sequence represents a human disease related nucleotide kinase
CC  1 (DRNK-1) DNA sequence first identified in Incyte Clone 56821 from the
CC  fibroblast cDNA library (FIBRNOR 01). The DRNK-1 DNA sequence is useful
CC  for the production of the corresponding recombinant enzyme. The invention
CC  provides DRNK enzymes, which are deoxyguanosine kinases, which catalyse
CC  the transfer of a terminal phosphate from adenosine triphosphate (ATP) or
CC  guanine triphosphate (GTP) to guanosine or guanine in the regulation of
CC  cellular levels of GTP and its corresponding nucleoside triphosphate. As
CC  GTP levels are known to control the activity of certain oncogenic
CC  proteins e.g. p21ras, a protein involved in cell proliferation and
CC  oncogenesis, suppression of the enzyme activity causes high ratios of
CC  GTP:GDP, promoting oncogenesis. Therefore, diseases (e.g. cancers, immune
CC  disorders and neurological dysfunction) related to this lack of activity
CC  may be prevented or treated with the recombinant enzyme, or by gene
CC  therapy based strategies. Anti-sense constructs of the DRNK encoding
CC  nucleic acids may also be used for inhibition of over-expression of the
CC  enzyme
XX
XX  Sequence 958 BP; 243 A; 254 C; 228 G; 231 T; 0 U; 2 Other;
SQ

```

```

Query Match 32.4%; Score 28.2; DB 2; Length 958;
Best Local Similarity 60.8%; Pred. No. 4.2;
Matches 45; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY      4  GAACCATCATTAATGAAGTGAATTTTCTGGCCTGAGACTTGAGGGAGGCAAGA 63
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      929  GAAAGNCGTTGGAGACACACATTCTTCTAGCTTCAGAAAGTCAGGAGCCCATCA 870
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      64  CACTCTGACACCA 77
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      869  CAGCCTGAACATCA 856
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
ADD48661
ID  ADD48661 standard; DNA; 2716 BP.
XX
XX  ADD48661;
XX
XX  29-JAN-2004 (first entry)
XX
XX  Rat gene L43592, SEQ ID NO 14367.
DE  Rat; ds; gene; pain; neuronal tissue; gene therapy;
KM  spinal segmental nerve injury; chronic constriction injury; CCI;
KM  spared nerve injury; SNI; Chung.
XX
XX  Rattus norvegicus.
OS
XX  WO2003016475-A2.
XX
XX  27-FEB-2003.
XX
XX  14-AUG-2002; 2002MO-US025765.
XX
XX  14-AUG-2001; 2001US-0312147P.
XX  01-NOV-2001; 2001US-0346382P.
XX  26-NOV-2001; 2001US-0333347P.
XX
XX  (GEHO) GEN HOSPITAL CORP.
XX  (FARB) BAYER AG.
XX
XX  Woolf C, D'urso D, Befort K, Costigan M;
PI  WPI; 2003-268312/26.
DR  GENBANK; L43592.
XX
XX  New composition comprising two or more isolated polypeptides, useful for
PT  preparing a medicament for treating pain in an animal.
XX
XX  Claim 1; Page; 1017pp; English.
XX
XX  The invention discloses a composition comprising two or more isolated rat
CC  or human polynucleotides or a polynucleotide which represents a fragment,
CC  derivative or allelic variation of the nucleic acid sequence. Also
CC  claimed are a vector comprising the novel polynucleotide, a host cell
CC  comprising the vector, a method for identifying a nucleotide sequence
CC  which is differentially regulated in an animal subjected to pain and a
CC  kit to perform the method, an array, a method for identifying an agent
CC  that increases or decreases the expression of the polynucleotide sequence
CC  that is differentially expressed in neuronal tissue of a first animal
CC  subjected to pain, a method for identifying a compound which regulates
CC  the expression of a polynucleotide sequence which is differentially
CC  expressed in an animal subjected to pain, a method for identifying a
CC  compound that regulates the activity of one or more of the
CC  polynucleotides, a method for producing a pharmaceutical composition, a
CC  method for identifying a compound or small molecule that regulates the
CC  activity in an animal of one or more of the polypeptides given in the
CC  specification, a method for identifying a compound useful in treating
CC  pain and a pharmaceutical composition comprising the one or more
CC  polypeptides or their antibodies. The polynucleotide or the compound that
CC  modulates its activity is useful for preparing a medicament for treating

```







GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 12, 2004, 17:39:25 ; Search time 6.47321 Seconds

(without alignments)  
7458.540 Million cell updates/sec

Title: US-09-966-880A-11

Perfect score: 87

Sequence: 1 agaaacaccatcaatga.....ctggacacacactatggacag 87

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:\*

1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*

2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*

3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq:\*

4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq:\*

5: /cgn2\_6/ptodata/2/ina/PCTUS.COMB.seq:\*

6: /cgn2\_6/ptodata/2/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	28.2	32.4	958	1	US-08-879-561-2
C 2	27.6	31.7	587	4	US-09-833-381-916
C 3	27.6	31.7	843	2	US-08-458-657-1
C 4	27.6	31.7	843	3	US-08-860-995-1
C 5	27.6	31.7	843	5	PCT-US95-00532A-1
C 6	27.6	31.7	1003	4	US-09-833-381-917
C 7	25.2	29.0	5448	4	US-09-620-312D-246
C 8	25.2	29.0	5532	4	US-09-620-312D-245
C 9	25.2	28.7	34185	4	US-09-545-481-3
C 10	24.6	28.3	786431	4	US-09-751-389-3
C 11	24.4	28.0	492	3	US-08-998-416-41
C 12	24.4	28.0	1631	3	US-08-895-590-43
C 13	24.2	27.8	2431	4	US-09-023-655-810
C 14	24.2	27.8	2546	4	US-09-598-401C-95
C 15	24.2	27.8	3016	4	US-09-016-434-1110
C 16	24.2	27.8	3035	5	PCT-US94-09235-1
C 17	24.2	27.8	168575	4	US-09-426-290-1
C 18	24.2	27.6	546	4	US-09-621-976-2448
C 19	23.8	27.4	1555	4	US-09-227-357-25
C 20	23.8	27.4	2582	3	US-09-130-616-176
C 21	23.8	27.4	4892	4	US-09-620-312D-1065
C 22	23.8	27.4	5537	1	US-08-135-511-32
C 23	23.8	27.4	5537	1	US-08-483-852-9
C 24	23.8	27.4	5537	1	US-08-361-458-4
C 25	23.8	27.4	5537	1	US-08-477-953-9
C 26	23.8	27.4	5537	1	US-08-187-453-32
C 27	23.8	27.4	5537	1	US-08-562-985A-5

28	23.8	27.4	5537	2	US-08-477-952-9	Sequence 9, Appli
C 29	23.8	27.4	9573	4	US-09-220-133-168	Sequence 168, App
C 30	23.8	27.4	51952	3	US-08-947-823-1	Sequence 1, Appli
C 31	23.8	27.4	64467	4	US-09-803-671B-3	Sequence 3, Appli
C 32	23.6	27.1	128779	4	US-09-497-855A-38	Sequence 38, Appli
C 33	23.6	26.9	551	3	US-09-040-984-25	Sequence 29, Appli
C 34	23.4	26.9	551	4	US-09-123-912-29	Sequence 29, Appli
C 35	23.4	26.9	551	4	US-09-643-597-29	Sequence 29, Appli
C 36	23.4	26.9	551	4	US-09-480-884A-29	Sequence 29, Appli
C 37	23.4	26.9	551	4	US-09-542-615A-29	Sequence 29, Appli
C 38	23.4	26.9	551	4	US-09-606-421B-29	Sequence 29, Appli
C 39	23.4	26.9	551	4	US-09-221-107-29	Sequence 29, Appli
C 40	23.4	26.9	4693	4	US-09-462-561B-8	Sequence 8, Appli
C 41	23.4	26.9	269223	4	US-09-596-002-41	Sequence 41, Appli
C 42	23.2	26.7	660	3	US-09-248-528-14	Sequence 14, Appli
C 43	23.2	26.7	660	3	US-09-549-108-14	Sequence 14, Appli
C 44	23.2	26.7	660	3	US-09-549-111-14	Sequence 14, Appli
C 45	23.2	26.7	660	3	US-09-549-106-14	Sequence 14, Appli

## ALIGNMENTS

RESULT 1  
US-08-879-561-2/c  
Sequence 2, Application US/08879561  
Patent No. 5817482  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Hawkin, Phillip R.  
APPLICANT: Guegler, Karl J.  
APPLICANT: Corley, Neil C.  
TITLE OF INVENTION: DISEASE RELATED NUCLEOTIDE KINASES  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/879,561  
FILING DATE: Herewith  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0325 US  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 958 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: FIBRONOT1  
CLONE: 56821  
US-08-879-561-2



Query Match 32.4%; Score 28.2; DB 1; Length 958;  
Best Local Similarity 60.8%; Pred. No. 0.44; 29; Indels 0; Gaps 0;  
Matches 45; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 4 GAACCATCATTAATTAAGTGAATTTTCTGGCTGAGACTTGACGAGGAGGCAAGAAGA 63  
DB 929 GAAAGTGTGGAGAGACACACATTTTCTAGCTTCAGAAAGTCAGGAGCCCAATCA 870

QY 64 CACTCTGACACCA 77  
DB 869 CAGCTGACATCA 856

RESULT 2  
US-09-833-381-916  
Sequence 916; Application US/09833381  
Patent No. 6672186

GENERAL INFORMATION:  
APPLICANT: ROBISON, Keith E.  
TITLE OF INVENTION: NO. 6672186el Nucleic Acid and Protein Homologs  
FILE REFERENCE: 5800-119  
CURRENT FILING DATE: 2001-04-11  
PRIOR APPLICATION NUMBER: 09/516,448  
PRIOR FILING DATE: 2000-02-29  
NUMBER OF SEQ ID NOS: 2050  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO: 916  
LENGTH: 587  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-833-381-916

Query Match 31.7%; Score 27.6; DB 4; Length 587;  
Best Local Similarity 60.8%; Pred. No. 0.62; 29; Indels 0; Gaps 0;  
Matches 45; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 4 GAACCATCATTAATTAAGTGAATTTTCTGGCTGAGACTTGACGAGGAGGCAAGAAGA 63  
DB 132 GAAAGTGTGGAGAGACACACATTTTCTAGCTTCAGAAAGTCAGGAGCCCAATCA 191

QY 64 CACTCTGACACCA 77  
DB 192 CAGCTGACATCA 205

RESULT 3  
US-08-458-657-1/c  
Sequence 1; Application US/08458657  
Patent No. 5914258  
GENERAL INFORMATION:  
APPLICANT: WEI, ET AL.  
TITLE OF INVENTION: Human Deoxycytidine Kinase 2  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CECCHI, STEWART & OLSTEIN  
STREET: 6 BECKER FARM ROAD  
CITY: ROSELAND  
STATE: NEW JERSEY  
COUNTRY: USA  
ZIP: 07068  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH DISKETTE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WORD PERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/458,657  
FILING DATE: June 2, 1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/00532

FILING DATE: 13 JAN 95  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: PERRARO, GREGORY D.

REGISTRATION NUMBER: 36,134  
REFERENCE/DOCKET NUMBER: 325800-371  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-994-1700  
TELEFAX: 201-994-1744

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:  
LENGTH: 843 BASE PAIRS  
TYPE: NUCLEIC ACID  
STRANDEDNESS: SINGLE  
TOPOLOGY: LINEAR  
MOLECULE TYPE: CDNA  
US-08-458-657-1

Query Match 31.7%; Score 27.6; DB 2; Length 843;  
Best Local Similarity 60.8%; Pred. No. 0.7; 29; Indels 0; Gaps 0;  
Matches 45; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 4 GAACCATCATTAATTAAGTGAATTTTCTGGCTGAGACTTGACGAGGAGGCAAGAAGA 63  
DB 690 GAAAGTGTGGAGAGACACACATTTTCTAGCTTCAGAAAGTCAGGAGCCCAATCA 631

QY 64 CACTCTGACACCA 77  
DB 630 CAGCTGACATCA 617

RESULT 4  
US-08-860-995-1/c  
Sequence 1; Application US/08860995  
Patent No. 6063376  
GENERAL INFORMATION:  
APPLICANT: WEI, ET AL.  
TITLE OF INVENTION: Human Deoxycytidine Kinase 2  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CECCHI, STEWART & OLSTEIN  
STREET: 6 BECKER FARM ROAD  
CITY: ROSELAND  
STATE: NEW JERSEY  
COUNTRY: USA  
ZIP: 07068  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 INCH DISKETTE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WORD PERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/860,995  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/00532  
FILING DATE: 13 JAN 95  
ATTORNEY/AGENT INFORMATION:  
NAME: PERRARO, GREGORY D.  
REGISTRATION NUMBER: 36,134  
REFERENCE/DOCKET NUMBER: 325800-246  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-994-1700  
TELEFAX: 201-994-1744  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 843 BASE PAIRS  
TYPE: NUCLEIC ACID  
STRANDEDNESS: SINGLE  
TOPOLOGY: LINEAR  
MOLECULE TYPE: CDNA



US-08-860-995-1

Query Match 31.7%; Score 27.6; DB 3; Length 843;  
Best Local Similarity 60.8%; Pred. No. 0.7;  
Matches 45; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 4 GAACCATCATTAATTGAAGTGAATTTTCTGCGCTGAGACTTGACAGGAGGCAAGAGA 63  
DB 690 GAAAGGTGGTTGGAGACACAACTTTTCTAGCTTGAGAAAGTCAGGAGGCCCAATCA 631

QY 64 CACTCTGACACCA 77  
DB 630 CAGCCTGAACATCA 617

## RESULT 5

PCT-US95-00532A-1/c  
Sequence 1, Application PC/TUS9500532A  
GENERAL INFORMATION:

APPLICANT: WEI, ET AL.  
TITLE OF INVENTION: Human Deoxycytidine Kinase 2  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,  
ADDRESSEE: CECCHI, STEWART & OLSTEIN  
STREET: 6 BECKER FARM ROAD  
CITY: ROSELAND  
STATE: NEW JERSEY  
COUNTRY: USA  
ZIP: 07068

## COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 INCH DISKETTE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WORD PERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/00532A  
FILING DATE: 13 JAN 95  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
CLASSIFICATION:

## ATTORNEY/AGENT INFORMATION:

NAME: FERRARO, GREGORY D.  
REGISTRATION NUMBER: 36,134  
REFERENCE/DOCKET NUMBER: 325800-246  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-994-1700  
TELEFAX: 201-994-1744

## INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:  
LENGTH: 843 BASE PAIRS  
TYPE: NUCLEIC ACID  
STRANDEDNESS: SINGLE  
TOPOLOGY: LINEAR  
MOLECULE TYPE: CDNA  
PCT-US95-00532A-1

Query Match 31.7%; Score 27.6; DB 5; Length 843;  
Best Local Similarity 60.8%; Pred. No. 0.7;  
Matches 45; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 4 GAACCATCATTAATTGAAGTGAATTTTCTGCGCTGAGACTTGACAGGAGGCAAGAGA 63  
DB 690 GAAAGGTGGTTGGAGACACAACTTTTCTAGCTTGAGAAAGTCAGGAGGCCCAATCA 631

QY 64 CACTCTGACACCA 77  
DB 630 CAGCCTGAACATCA 617

## RESULT 6

US-09-833-381-917

Sequence 917, Application US/09833381

Patent No. 6672186

GENERAL INFORMATION:

APPLICANT: Robison, Keith E.

TITLE OF INVENTION: No. 6672186el Nucleic Acid and Protein Homologs

FILE REFERENCE: 5800-119

CURRENT APPLICATION NUMBER: US/09/833,381

CURRENT FILING DATE: 2001-04-11

PRIOR APPLICATION NUMBER: 09/516,448

PRIOR FILING DATE: 2000-02-29

NUMBER OF SEQ ID NOS: 2050

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 917

LENGTH: 1003

TYPE: DNA

ORGANISM: Homo sapiens

US-09-833-381-917

Query Match 31.7%; Score 27.6; DB 4; Length 1003;  
Best Local Similarity 60.8%; Pred. No. 0.75; 29; Indels 0; Gaps 0;  
Matches 45; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 4 GAACCATCATTAATTGAAGTGAATTTTCTGCGCTGAGACTTGACAGGAGGCAAGAGA 63  
DB 178 GAAAGGTGGTTGGAGACACAACTTTTCTAGCTTGAGAAAGTCAGGAGGCCCAATCA 237

QY 64 CACTCTGACACCA 77  
DB 238 CAGCCTGAACATCA 251

## RESULT 7

US-09-620-312D-246

Sequence 246, Application US/09620312D

Patent No. 6569662

GENERAL INFORMATION:

APPLICANT: Tang, Y. Tom

APPLICANT: Liu, Chenghua

APPLICANT: Asundi, Vinod

APPLICANT: Zhang, Jie

APPLICANT: Ren, Feiyun

APPLICANT: Chen, Rui-hong

APPLICANT: Zhao, Qing A.

APPLICANT: Wehrman, Tom

APPLICANT: Xue, Aldong J.

APPLICANT: Yang, Yonhong

APPLICANT: Wang, Jian-Rui

APPLICANT: Zhou, Ping

APPLICANT: Ma, Yungling

APPLICANT: Wang, Dunrui

APPLICANT: Wang, Zhiwei

APPLICANT: John Tillinghast

APPLICANT: Drmanac, Radose T

TITLE OF INVENTION: No. 6569662el Nucleic Acids and

FILE REFERENCE: 784C1P2B

CURRENT APPLICATION NUMBER: US/09/620,312D

CURRENT FILING DATE: 2000-07-19

PRIOR APPLICATION NUMBER: 09/552,317

PRIOR FILING DATE: 2000-04-25

PRIOR APPLICATION NUMBER: 09/488,725

PRIOR FILING DATE: 2000-01-21

NUMBER OF SEQ ID NOS: 1105

SOFTWARE: pt\_FL\_genes Version 1.0

SEQ ID NO 246

LENGTH: 5448

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: (534)..(4178)

US-09-620-312D-246



Query Match 29.0%; Score 25.2; DB 4; Length 5448;  
Best Local Similarity 62.9%; Pred. No. 11;  
Matches 39; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 1 AGAGAACCATCATTAATTTGAAGTGAATTTTCTGGCCTGAGACTGACAGGAGGCAAGA 60  
DB 1910 AGATATAGACTTAATGGAAGAAATATATGCTGAGAGCTTCCAGAAAGGCAAAA 1969

QY 61 AG 62  
DB 1970 AG 1971

RESULT 8  
US-09-620-312D-245  
Sequence 245, Application US/09620312D

Patent No. 6569662  
GENERAL INFORMATION:  
APPLICANT: Tang, Y. Tom  
APPLICANT: Liu, Chenghua  
APPLICANT: Asundi, Vinod  
APPLICANT: Zhang, Jie  
APPLICANT: Ren, Feiyan  
APPLICANT: Chen, Rui-hong  
APPLICANT: Zhao, Qing A.  
APPLICANT: Wehrman, Tom  
APPLICANT: Xue, Aidong J.  
APPLICANT: Yang, Yonshong  
APPLICANT: Wang, Jian-Rui  
APPLICANT: Zhou, Ping  
APPLICANT: Ma, Yunding  
APPLICANT: Wang, Dunrui  
APPLICANT: Wang, Zhiwei  
APPLICANT: John Tillinghast  
APPLICANT: Drmanac, Radote T.  
TITLE OF INVENTION: No. 6569662el Nucleic Acids and  
TITLE OF INVENTION: Polypeptides  
FILE REFERENCE: 784CIP2B  
CURRENT APPLICATION NUMBER: US/09/620,312D  
CURRENT FILING DATE: 2000-07-19  
PRIOR APPLICATION NUMBER: 09/552,317  
PRIOR FILING DATE: 2000-04-25  
PRIOR APPLICATION NUMBER: 09/488,725  
PRIOR FILING DATE: 2000-01-21  
NUMBER OF SEQ ID NOS: 1105  
SOFTWARE: pt\_Fl\_genes Version 1.0  
SEQ ID NO 245  
LENGTH: 5532  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (534)..(4262)  
US-09-620-312D-245

Query Match 29.0%; Score 25.2; DB 4; Length 5532;  
Best Local Similarity 62.9%; Pred. No. 11;  
Matches 39; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 1 AGAGAACCATCATTAATTTGAAGTGAATTTTCTGGCCTGAGACTGACAGGAGGCAAGA 60  
DB 1910 AGATATAGACTTAATGGAAGAAATATATGCTGAGAGCTTCCAGAAAGGCAAAA 1969

QY 61 AG 62  
DB 1970 AG 1971

RESULT 9  
US-09-545-481-3  
Sequence 3, Application US/09545481  
Patent No. 6451319

GENERAL INFORMATION:  
APPLICANT: Chiang, H.  
APPLICANT: Cochran, Mark D.  
TITLE OF INVENTION: No. 6451319el Recombinant And Mutant Adenoviruses  
FILE REFERENCE: SY0993K US  
CURRENT APPLICATION NUMBER: US/09/545,481  
CURRENT FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: 60/128,766  
PRIOR FILING DATE: 1999-04-09  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: Patent Ver. 2.1  
SEQ ID NO 3  
LENGTH: 34185  
TYPE: DNA  
ORGANISM: Bovine adenovirus type 1  
US-09-545-481-3

Query Match 28.7%; Score 25; DB 4; Length 34185;  
Best Local Similarity 56.8%; Pred. No. 25;  
Matches 46; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 6 ACCATCATTAATTTGAAGTGAATTTTCTGGCCTGAGACTGACAGGAGGCAAGAAGA 65  
DB 29304 AACATCTTATTTGAGTCGTGAATTTGTGACATCACTTACATTCACCAACAC 29363  
QY 66 CTCGTGACACCACTATGACA 86  
DB 29364 CACTCGATGACACTTATACA 29384

RESULT 10  
US-09-751-389-3/C

Sequence 3, Application US/09751389  
Patent No. 6630334  
GENERAL INFORMATION:  
APPLICANT: GUEGLER, Karl et al  
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
TITLE OF INVENTION: THEREOF  
FILE REFERENCE: CL001067  
CURRENT APPLICATION NUMBER: US/09/751,389  
CURRENT FILING DATE: 2001-01-02  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 3  
LENGTH: 786431  
TYPE: DNA  
ORGANISM: Human  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)..(786431)  
OTHER INFORMATION: n = A,T,C or G  
US-09-751-389-3

Query Match 28.3%; Score 24.6; DB 4; Length 786431;  
Best Local Similarity 59.2%; Pred. No. 1e+02;  
Matches 42; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 15 AATTGAAGTGAATTTTCTGGCCTGAGACTTGCAGAGGAGGCAAGACACTTGTGACA 74  
DB 617689 AGTTTAAAGAGATGTTCTGGATTAGTTTACTACAGATACATTAATCATCTTGGCA 617630  
QY 75 CCACTATGAC 85  
DB 617629 GCACTATCTAC 617619

RESULT 11  
US-08-998-416-41/c  
Sequence 41, Application US/08998416  
Patent No. 6239264  
GENERAL INFORMATION:  
APPLICANT: Philippsen, Peter



APPLICANT: Pohlmann, Rainer  
APPLICANT: Steiner, Sabine  
APPLICANT: Mohr, Christine  
APPLICANT: Wendland, Jurgen  
APPLICANT: Knechtel, Philipp  
APPLICANT: Reibischung, Corinne  
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSEYP11  
TITLE OF INVENTION: AND USES THEREOF  
NUMBER OF SEQUENCES: 1152  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 6239264artis Corporation  
STREET: 3054 Cornwallis Road  
CITY: Research Triangle Park  
STATE: No. 6239264th Carolina  
COUNTRY: USA  
ZIP: 27709  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/998,416  
FILING DATE: 24-DEC-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: CH 0016/97  
FILING DATE: 31-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGCL976  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8587  
TELEFAX: 919-541-8689  
INFORMATION FOR SEQ ID NO: 41:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 492 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE: PAG1010UP  
ORGANISM: PAG1010UP  
US-08-998-416-41

Query Match 28.0%; Score 24.4; DB 3; Length 492;  
Best Local Similarity 68.0%; Pred. No. 9.1; Indels 16; Gaps 0;  
Matches 34; Conservative 0; Mismatches 16; Indels 16; Gaps 0;

Db 21 AGTGAATTTTCTGCGCTGAGACTTGACGAGGAGGCAAGACACTCTG 70  
461 ATTGAATTTTCTGCTGAGGCTTCACTGCTGGGATGAAGGCTCG 412

RESULT 12  
US-08-895-590-43/C  
Sequence 43, Application US/08895590  
Patent No. 6207410  
GENERAL INFORMATION:  
APPLICANT: Hall, Linda M.  
APPLICANT: Ren, Dejian  
APPLICANT: Zheng, Wei  
APPLICANT: Dubald, Manuel Marcel Paul  
TITLE OF INVENTION: Genes Encoding an Insect Calcium Channel  
NUMBER OF SEQUENCES: 101  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BURNS, DONALD, SWECKER & MATHIS, LLP  
STREET: 699 Prince Street  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22314-3187

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/895,590  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/374,888  
FILING DATE: 19-JAN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm M.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 022650-263  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-836-6620  
TELEFAX: 703-836-2021  
INFORMATION FOR SEQ ID NO: 43:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1631 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 33..1631  
US-08-895-590-43

Query Match 28.0%; Score 24.4; DB 3; Length 1631;  
Best Local Similarity 60.6%; Pred. No. 14; Indels 26; Gaps 0;  
Matches 40; Conservative 0; Mismatches 26; Indels 26; Gaps 0;

Db 15 AATTGAAGTGAATTTTCTGCGCTGAGACTTGACGAGGAGGCAAGACACTCTGACA 74  
1084 ATTGAAGGAGAGGTTTTCGCTCGACTTCCGACGCGATTGATCAAGGCTGTAA 1025

QY 75 CCACTA 80  
Db 1024 ACCTTA 1019

RESULT 13  
US-09-023-655-810/C  
Sequence 810, Application US/09023655  
Patent No. 6607879  
GENERAL INFORMATION:  
APPLICANT: Cocks, Benjamin G.  
APPLICANT: Susan G. Stuart  
APPLICANT: Jeffrey J. Sellhammer  
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE  
TITLE OF INVENTION: EXPRESSION  
NUMBER OF SEQUENCES: 1508  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/023,655  
FILING DATE: HEREWITH  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:



FILED DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Zeller, Karen J.  
REGISTRATION NUMBER: 37,071  
REFERENCE/DOCKET NUMBER: PA-0001 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 855-0555  
TELEFAX: (650) 845-4166  
INFORMATION FOR SEQ ID NO: 810:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2431 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: BRAIT102  
CLONE: 757560  
US-09-023-655-810

Query Match 27.8%; Score 24.2; DB 4; Length 2431;  
Best Local Similarity 66.0%; Pred. No. 19;  
Matches 35; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 8 CATCATTAATGAGTGAATTTTCTGGCCTGAGACTTGCGAGGAGCAAGA 60  
DB 616 CATCATTAATGAGTGAATTTTCTGGCCTGAGACTTGCGAGGAGCAAGA 564

RESULT 14  
US-09-598-401C-95/C  
Sequence 95, Application US/09598401C  
Patent No. 6596925  
GENERAL INFORMATION:  
APPLICANT: Perera, J. Ranjan  
APPLICANT: Eagleton, Clare  
APPLICANT: Rice, Stephen J.  
TITLE OF INVENTION: Compositions and Methods for the  
FILE OF INVENTION: Modification of Gene Expression  
FILE REFERENCE: 11000.1036C2  
CURRENT APPLICATION NUMBER: US/09/598,401C  
CURRENT FILING DATE: 2000-06-20  
PRIOR APPLICATION NUMBER: U.S. No. 6596925 60/146,591  
PRIOR FILING DATE: 1999-07-30  
PRIOR APPLICATION NUMBER: PCT/NZ00/00018  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: U.S. No. 6596925 09/276,599  
PRIOR FILING DATE: 1999-03-25  
NUMBER OF SEQ ID NOS: 120  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 95  
LENGTH: 2546  
TYPE: DNA  
ORGANISM: Pinus radiata  
US-09-598-401C-95

Query Match 27.8%; Score 24.2; DB 4; Length 2546;  
Best Local Similarity 57.1%; Pred. No. 20;  
Matches 44; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 5 AACCATCAATTAATGAAGTGAATTTTCTGGCCTGAGACTTGCGAGGAGCAAGA 64  
DB 2472 AACCATCAATGAAGTGAATTTTCTGGCCTGAGACTTGCGAGGAGCAAGA 2413

QY 65 ACTCTGACACCACTAT 81  
DB 2412 ACTCGAAGCAATTTT 2396

RESULT 15  
US-09-016-434-1110  
Sequence 1110, Application US/09016434  
Patent No. 6500938

GENERAL INFORMATION:  
APPLICANT: Janice Au-Young  
APPLICANT: Jeffrey J. Seilhamer  
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING  
TITLE OF INVENTION: PATHWAY GENE EXPRESSION  
NUMBER OF SEQUENCES: 1490  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/016,434  
FILING DATE: HEREWITH  
CLASSIFICATION:  
APPLICATION DATA:  
FILING DATE:  
APPLICATION NUMBER:  
ATTORNEY/AGENT INFORMATION:  
NAME: Zeller, Karen J.  
REGISTRATION NUMBER: 37,071  
REFERENCE/DOCKET NUMBER: PA-0002 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 855-0555  
TELEFAX: (650) 845-4166  
INFORMATION FOR SEQ ID NO: 1110:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3016 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GENBANK  
CLONE: 91321593  
US-09-016-434-1110

Query Match 27.8%; Score 24.2; DB 4; Length 3016;  
Best Local Similarity 66.0%; Pred. No. 21;  
Matches 35; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 1 AGAAGACATCAATTAATGAAGTGAATTTTCTGGCCTGAGACTTGCGAGGGA 53  
DB 1294 AGAAGACATCAATTAATGAAGTGAATTTTCTGGCCTGAGACTTGCGAGGGA 1346

Search completed: March 13, 2004, 00:51:06  
Job time: 8.47321 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model1

Run on: March 12, 2004, 21:22:51 ; Search time 30.53 Seconds

(Without alignments)  
10491.276 Million cell updates/sec

Title: US-09-966-880A-11

Perfect score: 87

Sequence: 1 agagaaccatcatatga.....ctggacacatgatgacag 87

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2412557 seqs, 1840798884 residues

Total number of hits satisfying chosen parameters: 4865114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Published Applications NA:\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PTI\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US09\_PUB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
17: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
18: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	87	100.0	87	US-09-966-880A-11	Sequence 11, Appl
2	87	100.0	2818	US-09-966-880A-7	Sequence 9, Appl
3	87	100.0	5514	US-09-966-880A-9	Sequence 35, Appl
4	87	100.0	11204	US-09-966-880A-35	Sequence 1, Appl
5	41	47.1	2440	US-09-966-880A-1	Sequence 1014, Ap
6	30	34.5	328	US-09-728-446-1014	Sequence 164874,
7	28	32.2	809	US-10-027-632-164874	Sequence 7046, Ap
8	27.6	31.7	327	US-10-085-783A-7046	Sequence 5294, Ap
9	27.6	31.7	333	US-10-242-535A-7046	Sequence 5294, Ap
10	27.6	31.7	333	US-10-040-862-5294	Sequence 5294, Ap
11	27.6	31.7	333	US-10-057-475B-5294	Sequence 5294, Ap
12	27.6	31.7	333	US-10-154-864B-5294	Sequence 5294, Ap
13	27.6	31.7	587	US-09-833-381-916	Sequence 916, App
14	27.6	31.7	587	US-09-833-381-916	Sequence 917, App
15	27.6	31.7	1003	US-09-833-381-917	Sequence 917, App

16	27.4	31.5	947	10	US-09-974-879-101	Sequence 101, App
17	27.4	31.5	947	11	US-09-305-736-101	Sequence 101, App
18	27.4	31.5	947	11	US-09-818-683-101	Sequence 101, App
19	27.4	31.5	947	12	US-10-621-401-101	Sequence 101, App
20	27.2	31.3	3119	14	US-10-198-846-13354	Sequence 13354, A
21	26.8	30.8	130320	15	US-10-408-168-1	Sequence 1, Appl
22	26.6	30.6	1706	12	US-10-424-599-44672	Sequence 44672, A
23	26.4	30.3	593	15	US-10-027-632-253079	Sequence 253079, A
24	26.4	30.3	593	15	US-10-027-632-253080	Sequence 253080, A
25	26.4	30.3	593	15	US-10-027-632-253081	Sequence 253081, A
26	26.4	30.3	609	15	US-10-027-632-253082	Sequence 10277, A
27	26.4	30.3	1101	15	US-10-027-632-31144	Sequence 31144, A
28	26.4	30.3	1164	15	US-10-027-632-30459	Sequence 30459, A
29	26.2	30.1	630	15	US-10-027-632-96683	Sequence 96683, A
30	26.2	30.1	630	15	US-10-027-632-308335	Sequence 308335, A
31	26.2	30.1	814	15	US-10-027-632-173408	Sequence 173408, A
32	26.2	30.1	32154	9	US-09-764-877-3433	Sequence 3433, Ap
33	26.2	30.1	32154	9	US-10-242-515-3433	Sequence 3433, Ap
34	26.2	30.1	32154	15	US-10-242-515-3433	Sequence 31, Appl
35	26.2	29.9	26	9	US-09-966-880A-31	Sequence 703, Appl
36	26.2	29.9	412	14	US-10-101-510-703	Sequence 62, Appl
37	26.2	29.9	417	14	US-10-082-828A-62	Sequence 8915, Ap
38	26.2	29.9	435	9	US-09-796-692-8915	Sequence 8915, Ap
39	26.2	29.9	435	14	US-10-040-862-8915	Sequence 8915, Ap
40	26.2	29.9	435	15	US-10-057-475B-8915	Sequence 8915, Ap
41	26.2	29.9	580	15	US-10-154-864B-8915	Sequence 279917, A
42	26.2	29.9	779	15	US-10-027-632-13154	Sequence 13154, A
43	26.2	29.9	887	15	US-10-103-959A-1	Sequence 1, Appl
44	26.2	29.9	985	15	US-10-103-959A-4	Sequence 4, Appl
45	26.2	29.9	1328	14	US-10-082-828A-63	Sequence 63, Appl

## ALIGNMENTS

RESULT 1  
US-09-966-880A-11  
Sequence 11, Application US/09966880A  
Patent No. US20020164743A1  
GENERAL INFORMATION:  
APPLICANT: Honjo, Tasuku  
TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
FILE REFERENCE: 06501-088001  
CURRENT APPLICATION NUMBER: US/09/966, 880A  
CURRENT FILING DATE: 2001-09-28  
PRIOR APPLICATION NUMBER: PCT/JPO0/01918  
PRIOR FILING DATE: 2000-03-28  
PRIOR APPLICATION NUMBER: JP 11-371382  
PRIOR FILING DATE: 1999-12-27  
PRIOR APPLICATION NUMBER: JP 11-178999  
PRIOR FILING DATE: 1999-06-24  
PRIOR APPLICATION NUMBER: JP 11-87192  
PRIOR FILING DATE: 1999-03-29  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 11  
LENGTH: 87  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-966-880A-11

Query Match 100.0%; Score 87; DB 9; Length 87;  
Best Local Similarity 100.0%; Pred. No. 7.6e-23;  
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAACAATCAATTAATTAAGTGAATTTTCTGGCTTGACCTTGACGAGGACGACGAGA 60  
DB 1 AGAACAATCAATTAATTAAGTGAATTTTCTGGCTTGACCTTGACGAGGACGACGAGA 60  
QY 61 AGACACTTGGACACCACTATGACGAG 87  
DB 61 AGACACTTGGACACCACTATGACGAG 87



```
RESULT 2
US-09-966-880A-7
; Sequence 7, Application US/09966880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; APPLICANT: Muramatsu, Masamichi
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966,880A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 2818
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (80)...(673)
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1)...(79)
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: (677)...(2818)
US-09-966-880A-7

Query Match
Best Local Similarity 100.0%; Score 87; DB 9; Length 2818;
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAAGACCATCATTAATTGAAGTGAATTTTCTGCGCTGAGACTTGACGGAGGCAAGA 60
DB 1 AGAAGACCATCATTAATTGAAGTGAATTTTCTGCGCTGAGACTTGACGGAGGCAAGA 60

QY 61 AGACACTCTGACACCACTATGACAG 87
DB 61 AGACACTCTGACACCACTATGACAG 87

RESULT 3
US-09-966-880A-9
; Sequence 9, Application US/09966880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; APPLICANT: Muramatsu, Masamichi
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966,880A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
```

```
LENGTH: 5514
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: Intron
; LOCATION: (1)...(1031)
; FEATURE:
; NAME/KEY: exon
; LOCATION: (1032)...(1118)
; FEATURE:
; NAME/KEY: Intron
; LOCATION: (1119)...(5514)
US-09-966-880A-9

Query Match
Best Local Similarity 100.0%; Score 87; DB 9; Length 5514;
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAAGACCATCATTAATTGAAGTGAATTTTCTGCGCTGAGACTTGACGGAGGCAAGA 60
DB 1032 AGAAGACCATCATTAATTGAAGTGAATTTTCTGCGCTGAGACTTGACGGAGGCAAGA 1091

QY 61 AGACACTCTGACACCACTATGACAG 87
DB 1092 AGACACTCTGACACCACTATGACAG 1118

RESULT 4
US-09-966-880A-35
; Sequence 35, Application US/09966880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; APPLICANT: Muramatsu, Masamichi
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966,880A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 35
; LENGTH: 11204
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-966-880A-35

Query Match
Best Local Similarity 100.0%; Score 87; DB 9; Length 11204;
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAAGACCATCATTAATTGAAGTGAATTTTCTGCGCTGAGACTTGACGGAGGCAAGA 60
DB 442 AGAAGACCATCATTAATTGAAGTGAATTTTCTGCGCTGAGACTTGACGGAGGCAAGA 501

QY 61 AGACACTCTGACACCACTATGACAG 87
DB 502 AGACACTCTGACACCACTATGACAG 528

RESULT 5
US-09-966-880A-1
; Sequence 1, Application US/09966880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
```



APPLICANT: Muramatsu, Masamichi  
TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
FILE REFERENCE: 06501-088001  
CURRENT APPLICATION NUMBER: US/09/966,880A  
CURRENT FILING DATE: 2001-09-28  
PRIOR APPLICATION NUMBER: PCT/JP00/01918  
PRIOR FILING DATE: 2000-03-28  
PRIOR APPLICATION NUMBER: JP 11-371382  
PRIOR FILING DATE: 1999-12-27  
PRIOR APPLICATION NUMBER: JP 11-178999  
PRIOR FILING DATE: 1999-06-24  
PRIOR APPLICATION NUMBER: JP 11-87192  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1  
LENGTH: 2440  
TYPE: DNA  
ORGANISM: Mus musculus  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (93)...(686)  
FEATURE:  
NAME/KEY: 5'UTR  
LOCATION: (1)...(92)  
FEATURE:  
NAME/KEY: 3'UTR  
LOCATION: (690)...(2440)  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(2440)  
OTHER INFORMATION: n = A,T,C or G  
US-09-966-880A-1

Query Match 47.1%; Score 41; DB 9; Length 2440;  
Best Local Similarity 75.3%; Pred. No. 5.3e-05;  
Matches 64; Conservative 0; Mismatches 20; Indels 1; Gaps 1;

QY 4 GAACCATCATTAATGAAGTGAATTTTCTGGCCTGAGACTTGAGGAGGCAAG-AAG 62  
DB 16 GAAGAGCCTTCTGGAAGCAAGCTTCTGGCTTAAGACTTGAAGGAGTCAAGAAAG 75  
QY 63 AACTCTGACACCCACTATGACAG 87  
DB 76 TCACGCTGAGACCGATATGACAG 100

RESULT 6  
US-09-728-446-1014  
Sequence 1014, Application US/09728446  
Patent No. US20020081668A1  
GENERAL INFORMATION:  
APPLICANT: Friedrich, Glenn  
APPLICANT: Zambrowicz, Brian  
TITLE OF INVENTION: No. US20020081668A1 Murine Polynucleotide Sequences  
TITLE OF INVENTION: and Mutant Cells and Mutant Animals Defined Thereby  
FILE REFERENCE: LEX-0101-USA  
CURRENT APPLICATION NUMBER: US/09/728,446  
CURRENT FILING DATE: 2000-11-30  
PRIOR APPLICATION NUMBER: US 60/168,270  
PRIOR FILING DATE: 1999-12-01  
NUMBER OF SEQ ID NOS: 1461  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1014  
LENGTH: 328  
TYPE: DNA  
ORGANISM: Mus musculus  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(328)  
OTHER INFORMATION: n = A,T,C or G  
US-09-728-446-1014

Query Match 34.5%; Score 30; DB 9; Length 328;  
Best Local Similarity 59.3%; Pred. No. 0.38;  
Matches 51; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 1 AGAAGACCATTAATGAAGTGAATTTTCTGGCCTGAGACTTGAGGAGGCAAGCA 60  
DB 103 AGAAGACCATTAATGAAGTGAATTTTCTGGCCTGAGACTTGAGGAGGCAAGCA 162  
QY 61 AGAAGACCATTAATGAAGTGAATTTTCTGGCCTGAGACTTGAGGAGGCAAGCA 86  
DB 163 AACGAGCAAGTGAAGTTCGGGAAA 188

RESULT 7  
US-10-027-632-164874  
Sequence 164874, Application US/10027632  
Publication No. US20030204075A9  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
POLYMORPHISMS IN THE HUMAN GENOME  
FILE REFERENCE: 108827.129  
CURRENT APPLICATION NUMBER: US/10/027,632  
CURRENT FILING DATE: 2002-04-30  
PRIOR APPLICATION NUMBER: US 60/218,006  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/198,676  
PRIOR FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: US 60/193,483  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: US 60/185,218  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/167,363  
PRIOR FILING DATE: 1999-11-23  
PRIOR APPLICATION NUMBER: US 60/156,358  
PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: US 60/146,002  
NUMBER OF SEQ ID NOS: 325720  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 164874  
LENGTH: 809  
TYPE: DNA  
ORGANISM: Human  
US-10-027-632-164874

Query Match 32.2%; Score 28; DB 15; Length 809;  
Best Local Similarity 63.2%; Pred. No. 2.9;  
Matches 43; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 10 TCATTAATGAAGTGAATTTTCTGGCCTGAGACTTGAGGAGGCAAGCAAGCACTCT 69  
DB 335 TGAGCACTTAATGAAGTGAATTTTCTGGCCTGAGACTTGAGGAGGCAAGCAAGCACT 394  
QY 70 GAGACCA 77  
DB 395 GAGTTCA 402  
RESULT 8  
US-10-085-783A-7046/C  
Sequence 7046, Application US/10085783A  
Publication No. US20040037841A1  
GENERAL INFORMATION:  
APPLICANT: ChondroGene Inc.  
APPLICANT: Liew, C.C.  
TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis  
FILE REFERENCE: 4231/2002  
CURRENT APPLICATION NUMBER: US/10/085,783A  
CURRENT FILING DATE: 2002-02-28  
PRIOR APPLICATION NUMBER: US 60/305,340  
PRIOR FILING DATE: 2001-07-13



PRIOR APPLICATION NUMBER: US 60/275,017  
PRIOR FILING DATE: 2001-03-12  
PRIOR APPLICATION NUMBER: US 60/271,955  
PRIOR FILING DATE: 2001-02-28  
NUMBER OF SEQ ID NOS: 58994  
SOFTWARE: Patent version 3.2  
SEQ ID NO 7046  
LENGTH: 327  
TYPE: DNA  
ORGANISM: Human  
US-10-085-783A-7046

Query Match 31.7%; Score 27.6; DB 12; Length 327;  
Best Local Similarity 60.8%; Pred. No. 3.1;  
Matches 45; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 4 GAACCATCATTAATGAGTGAATTTTCTGCGCTGAGACTTGACAGGAGGCAAGAGA 63  
DB 189 GAAAGGTGTTGGAGACACAACATTTTCTAGCTTCGAAAGTCAGGGAGCCAGATCA 130  
QY 64 CACTTGACACCA 77  
DB 129 CAGCTGAACATCA 116

RESULT 9  
US-10-242-535A-7046/C  
Sequence 7046, Application US/10242535A  
Publication No. US20040013663A1  
GENERAL INFORMATION:  
APPLICANT: ChondroGene Inc.

TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis  
FILE REFERENCE: 4231/2005  
CURRENT APPLICATION NUMBER: US/10/242,535A  
CURRENT FILING DATE: 2002-09-12  
PRIOR APPLICATION NUMBER: US 10/085,783  
PRIOR FILING DATE: 2002-02-28  
PRIOR APPLICATION NUMBER: US 60/305,340  
PRIOR FILING DATE: 2001-07-13  
PRIOR APPLICATION NUMBER: US 60/275,017  
PRIOR FILING DATE: 2001-03-12  
PRIOR APPLICATION NUMBER: US 60/271,955  
PRIOR FILING DATE: 2001-02-28  
NUMBER OF SEQ ID NOS: 58994  
SOFTWARE: Patent version 3.2  
SEQ ID NO 7046  
LENGTH: 327  
TYPE: DNA  
ORGANISM: Human  
US-10-242-535A-7046

Query Match 31.7%; Score 27.6; DB 15; Length 327;  
Best Local Similarity 60.8%; Pred. No. 3.1;  
Matches 45; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 4 GAACCATCATTAATGAGTGAATTTTCTGCGCTGAGACTTGACAGGAGGCAAGAGA 63  
DB 189 GAAAGGTGTTGGAGACACAACATTTTCTAGCTTCGAAAGTCAGGGAGCCAGATCA 130  
QY 64 CACTTGACACCA 77  
DB 129 CAGCTGAACATCA 116

RESULT 10  
US-09-796-692-5294/C  
Sequence 5294, Application US/09796692  
Publication No. US20020198362A1  
GENERAL INFORMATION:  
APPLICANT: Gaiger, Alexander  
APPLICANT: Algate, Paul A.  
APPLICANT: Mannion, Jane

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY  
FILE REFERENCE: 2077,001200  
CURRENT APPLICATION NUMBER: US/09/796,692  
PRIOR APPLICATION NUMBER: 60/186,126  
PRIOR FILING DATE: 2000-03-01  
PRIOR APPLICATION NUMBER: 60/190,479  
PRIOR FILING DATE: 2000-03-17  
PRIOR APPLICATION NUMBER: 60/200,545  
PRIOR FILING DATE: 2000-04-27  
PRIOR APPLICATION NUMBER: 60/200,303  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/200,779  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/200,999  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 60/202,084  
PRIOR FILING DATE: 2000-05-04  
PRIOR APPLICATION NUMBER: 60/206,201  
PRIOR FILING DATE: 2000-05-22  
PRIOR APPLICATION NUMBER: 60/218,950  
PRIOR FILING DATE: 2000-07-14  
PRIOR APPLICATION NUMBER: 60/222,903  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: 60/223,416  
PRIOR FILING DATE: 2000-08-04  
PRIOR APPLICATION NUMBER: 60/223,378  
PRIOR FILING DATE: 2000-08-07  
NUMBER OF SEQ ID NOS: 9597  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 5294  
LENGTH: 333  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (9)  
OTHER INFORMATION: n=A,T,C or G  
US-09-796-692-5294

Query Match 31.7%; Score 27.6; DB 9; Length 333;  
Best Local Similarity 60.8%; Pred. No. 3.1;  
Matches 45; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 4 GAACCATCATTAATGAGTGAATTTTCTGCGCTGAGACTTGACAGGAGGCAAGAGA 63  
DB 248 GAAAGGTGTTGGAGACACAACATTTTCTAGCTTCGAAAGTCAGGGAGCCAGATCA 189  
QY 64 CACTTGACACCA 77  
DB 188 CAGCTGAACATCA 175

RESULT 11  
US-10-040-862-5294/C  
Sequence 5294, Application US/10040862  
Publication No. US20030078396A1  
GENERAL INFORMATION:  
APPLICANT: Gaiger, Alexander  
APPLICANT: Mannion, Jane  
APPLICANT: Retler, Marc  
APPLICANT: Corixa Corporation  
TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy  
FILE REFERENCE: 014058-013520US  
CURRENT APPLICATION NUMBER: US/10/040,862  
CURRENT FILING DATE: 2001-11-06  
PRIOR APPLICATION NUMBER: US 60/186,126  
PRIOR FILING DATE: 2000-03-01  
PRIOR APPLICATION NUMBER: US 60/190,479  
PRIOR FILING DATE: 2000-03-17



```
;; PRIOR APPLICATION NUMBER: US 60/200,545
;; PRIOR FILING DATE: 2000-04-27
;; PRIOR APPLICATION NUMBER: US 60/200,303
;; PRIOR FILING DATE: 2000-04-28
;; PRIOR APPLICATION NUMBER: US 60/200,779
;; PRIOR FILING DATE: 2000-04-28
;; PRIOR APPLICATION NUMBER: US 60/200,999
;; PRIOR FILING DATE: 2000-05-01
;; PRIOR APPLICATION NUMBER: US 60/202,084
;; PRIOR FILING DATE: 2000-05-04
;; PRIOR APPLICATION NUMBER: US 60/206,201
;; PRIOR FILING DATE: 2000-05-22
;; PRIOR APPLICATION NUMBER: US 60/218,950
;; PRIOR FILING DATE: 2000-07-14
;; PRIOR APPLICATION NUMBER: US 60/222,903
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: US 60/223,416
;; PRIOR FILING DATE: 2000-08-04
;; PRIOR APPLICATION NUMBER: US 60/223,378
;; PRIOR FILING DATE: 2000-08-07
;; PRIOR APPLICATION NUMBER: US 09/796,692
;; PRIOR FILING DATE: 2001-03-01
;; NUMBER OF SEQ ID NOS: 10467
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 5294
;; LENGTH: 333
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: unsure
;; LOCATION: (9)
;; OTHER INFORMATION: n=A,T,C or G
US-10-040-862-5294

Query Match      31.7%; Score 27.6; DB 14; Length 333;
Best Local Similarity 60.8%; Pred. No. 3.1;
Matches 45; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY      4  GAACCATCATTAATGAGTGGATTTTCTGCGCTGAGACTTGAGGGAGGCAAGA 63
DB      248  GAAGGTGGTGGAGACACAACTTTCTAGCTTCAGAAAGTCAGGGAGCCAGATCA 189
QY      64  CACTCTGACACCA 77
DB      188  CAGCCTGAACATCA 175

RESULT 12
US-10-057-475B-5294/C
;; Sequence 5294, Application US/10057475B
;; Publication No. US20040002068A1
;; GENERAL INFORMATION:
;; APPLICANT: Gaiger, Alexander
;; APPLICANT: Algate, Paul A.
;; APPLICANT: Mannion, Jane
;; APPLICANT: Clapper, Jonathan David
;; APPLICANT: Wang, Aijun
;; APPLICANT: Ordenez, Nadia
;; APPLICANT: Carter, Lauren
;; APPLICANT: McNeill, Patricia Dianne
;; APPLICANT: Corixa Corporation
;; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
;; FILE REFERENCE: 014058-014402US
;; CURRENT APPLICATION NUMBER: US/10/057,475B
;; PRIOR APPLICATION NUMBER: US 60/186,126
;; PRIOR FILING DATE: 2000-03-01
;; PRIOR APPLICATION NUMBER: US 60/190,479
;; PRIOR FILING DATE: 2000-03-17
;; PRIOR APPLICATION NUMBER: US 60/200,545
;; PRIOR FILING DATE: 2000-04-27
;; PRIOR APPLICATION NUMBER: US 60/200,303
```

```
;; PRIOR FILING DATE: 2000-04-28
;; PRIOR APPLICATION NUMBER: US 60/200,779
;; PRIOR FILING DATE: 2000-04-28
;; PRIOR APPLICATION NUMBER: US 60/200,999
;; PRIOR FILING DATE: 2000-05-01
;; PRIOR APPLICATION NUMBER: US 60/202,084
;; PRIOR FILING DATE: 2000-05-04
;; PRIOR APPLICATION NUMBER: US 60/206,201
;; PRIOR FILING DATE: 2000-05-22
;; PRIOR APPLICATION NUMBER: US 60/218,950
;; PRIOR FILING DATE: 2000-07-14
;; PRIOR APPLICATION NUMBER: US 60/222,903
;; PRIOR FILING DATE: 2000-08-03
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 10979
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 5294
;; LENGTH: 333
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: misc feature
;; LOCATION: (1)...(333)
;; OTHER INFORMATION: n = G, A, C or T
US-10-057-475B-5294

Query Match      31.7%; Score 27.6; DB 15; Length 333;
Best Local Similarity 60.8%; Pred. No. 3.1;
Matches 45; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY      4  GAACCATCATTAATGAGTGGATTTTCTGCGCTGAGACTTGAGGGAGGCAAGA 63
DB      248  GAAGGTGGTGGAGACACAACTTTCTAGCTTCAGAAAGTCAGGGAGCCAGATCA 189
QY      64  CACTCTGACACCA 77
DB      188  CAGCCTGAACATCA 175

RESULT 13
US-10-154-884B-5294/C
;; Sequence 5294, Application US/10154884B
;; Publication No. US20040005561A1
;; GENERAL INFORMATION:
;; APPLICANT: Gaiger, Alexander
;; APPLICANT: Algate, Paul A.
;; APPLICANT: Mannion, Jane
;; APPLICANT: Retter, Marc W.
;; APPLICANT: Corixa Corporation
;; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
;; FILE REFERENCE: 014058-013521US
;; CURRENT APPLICATION NUMBER: US/10/154,884B
;; PRIOR APPLICATION NUMBER: US 60/186,126
;; PRIOR FILING DATE: 2000-03-01
;; PRIOR APPLICATION NUMBER: US 60/190,479
;; PRIOR FILING DATE: 2000-03-17
;; PRIOR APPLICATION NUMBER: US 60/200,545
;; PRIOR FILING DATE: 2000-04-27
;; PRIOR APPLICATION NUMBER: US 60/200,303
;; PRIOR FILING DATE: 2000-04-28
;; PRIOR APPLICATION NUMBER: US 60/200,779
;; PRIOR FILING DATE: 2000-04-28
;; PRIOR APPLICATION NUMBER: US 60/200,999
;; PRIOR FILING DATE: 2000-05-01
;; PRIOR APPLICATION NUMBER: US 60/202,084
;; PRIOR FILING DATE: 2000-05-04
;; PRIOR APPLICATION NUMBER: US 60/206,201
;; PRIOR FILING DATE: 2000-05-22
;; PRIOR APPLICATION NUMBER: US 60/218,950
;; PRIOR FILING DATE: 2000-07-14
;; PRIOR APPLICATION NUMBER: US 60/222,903
```



```
; PRIOR FILING DATE: 2000-08-03
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 11290
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5294
; LENGTH: 333
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(333)
; OTHER INFORMATION: n = g, a, c or t
US-10-154-884B-5294

Query Match
Best Local Similarity 31.7%; Score 27.6; DB 15; Length 333;
Matches 45; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

Db
4 GAACCATCATTAATTGAAGTGAAGATTTTCTGGCCTGAGACTTGCAAGGAGGCAAGAGA 63
QY 248 GAAAGTGTGGGAGACACAACTTTTCTAGCTTCAGAAAGTCAGGAGCCCGAGATCA 189
64 CACTGTGACACCA 77
Db 188 CAGCCTGAACATCA 175

RESULT 14
US-09-833-381-916
; Sequence 916, Application US/09833381
; Patent No. US20020132090A1
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: No. US20020132090A1 Nucleic Acid and Protein Homologs
; FILE REFERENCE: 5800-119
; CURRENT APPLICATION NUMBER: US/09/833,381
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 09/516,448
; NUMBER OF SEQ ID NOS: 2050
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 916
; LENGTH: 587
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-833-381-916

Query Match
Best Local Similarity 31.7%; Score 27.6; DB 9; Length 587;
Matches 45; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

Db
4 GAACCATCATTAATTGAAGTGAAGATTTTCTGGCCTGAGACTTGCAAGGAGGCAAGAGA 63
QY 132 GAAAGTGTGGGAGACACAACTTTTCTAGCTTCAGAAAGTCAGGAGCCCGAGATCA 191
64 CACTGTGACACCA 77
Db 192 CAGCCTGAACATCA 205

RESULT 15
US-09-833-381-917
; Sequence 917, Application US/09833381
; Patent No. US20020132090A1
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: No. US20020132090A1 Nucleic Acid and Protein Homologs
; FILE REFERENCE: 5800-119
; CURRENT APPLICATION NUMBER: US/09/833,381
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 09/516,448
; NUMBER OF SEQ ID NOS: 2050
```

```
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 917
; LENGTH: 1003
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-833-381-917

Query Match
Best Local Similarity 31.7%; Score 27.6; DB 9; Length 1003;
Matches 45; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

Db
4 GAACCATCATTAATTGAAGTGAAGATTTTCTGGCCTGAGACTTGCAAGGAGGCAAGAGA 63
QY 178 GAAAGTGTGGGAGACACAACTTTTCTAGCTTCAGAAAGTCAGGAGCCCGAGATCA 237
64 CACTGTGACACCA 77
Db 238 CAGCCTGAACATCA 251

Search completed: March 13, 2004, 05:30:54
Job time : 31.53 secs
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 12, 2004, 13:19:34 ; Search time 277.03 Seconds

(without alignments)  
9378.080 Million cell updates/sec

Title: US-09-966-880A-11

Perfect score: 87

Sequence: 1 agagacacacatcatatga.....ctggacacacacatgacag 87

Scoring table: IDENTITY\_NUC

Gapop 10.0, Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: EST:  
2: em\_estba:\*  
3: em\_esthum:\*  
4: em\_estin:\*  
5: em\_estm:\*  
6: em\_estov:\*  
7: em\_estpl:\*  
8: em\_estro:\*  
9: em\_hrc:\*  
10: gb\_est1:\*  
11: gb\_est2:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	87	100.0	535	14	CD707143 EST23670
2	87	100.0	693	12	BG757392 602711022
3	87	100.0	1201	13	BX402063 BX402063
4	85.4	98.2	820	12	BG757089 602715124

Result No.	Score	Query Match	Length	DB ID	Description
5	84	96.6	541	10	BF238155
6	84	96.6	743	12	BG686133
7	84	96.6	942	10	BF975166
8	71	81.6	953	13	B0065440
9	71	81.6	1052	13	B0055935
10	70.4	80.9	872	12	BG758510
11	39.2	45.1	1201	9	AL559877
12	38.8	44.6	853	13	BX464579
13	37	42.5	889	12	BG686876
14	31.6	36.3	930	28	CC198363
15	30.8	35.4	250	10	BE071397
16	30	34.5	192	29	CG530782
17	30	34.5	196	29	CG656390
18	30	34.5	204	29	CG545883
19	30	34.5	212	29	CG578867
20	30	34.5	590	28	BG620575
21	29.6	34.0	702	12	BG757541
22	29.4	33.8	329	28	AZ395639
23	29.4	33.8	508	28	AZ158608
24	29.2	33.6	272	13	BU732934
25	29	33.3	538	28	AQ779004
26	29	33.3	770	29	CG993610
27	28.8	33.1	612	29	CE326757
28	28.8	33.1	622	29	CE030990
29	28.8	33.1	661	14	CB419607
30	28.8	33.1	704	14	CA170646
31	28.8	33.1	736	38	AZ519070
32	28.8	33.1	1002	13	BQ709239
33	28.6	32.9	262	9	AV388203
34	28.6	32.9	375	28	AZ085135
35	28.6	32.9	896	14	CB627885
36	28.4	32.6	473	28	AQ859236
37	28.4	32.6	782	28	BE284924
38	28.4	32.6	947	28	CC537188
39	28.2	32.4	557	29	AX106106
40	28.2	32.4	628	29	BX128009
41	28	32.2	442	28	AQ276930
42	28	32.2	490	29	CG722866
43	28	32.2	599	14	CB243900
44	28	32.2	626	28	CC156720
45	28	32.2	803	29	CG093969

## ALIGNMENTS

RESULT 1  
CD707143 535 bp mRNA linear EST 25-JUN-2003  
DEFINITION EST23670 human nasopharynx Homo sapiens cDNA, mRNA sequence.  
CD707143  
VERSION CD707143.1 GI:32237773  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)

REFERENCE  
AUTHORS Liu,X.-Q., Zhou,Y., Zhang,L.-J., Xu,H., Chen,H.-K., Fan,Z.-G. and Zeng,Y.-X.  
TITLE Transcriptional Gene Expression Profile of Human Nasopharynx  
JOURNAL Unpublished (2003)  
COMMENT Contact: Yixin Zeng  
Cancer Center  
Sun Yat-sen University  
651 Dongfeng Road East, Guangzhou 510060, China  
Tel: 86-1380-9770-743  
Fax: 86-20-8775-4506  
Email: yxzeng@zsu.edu.cn.  
Location/Qualifiers  
1. 535  
/organism="Homo sapiens"  
/mol\_type="mRNA"

## FEATURES

source



```

/db_xref="taxon:9606"
/tissue_type="normal nasopharynx"
/clonue_lib="human nasopharynx"
/ncore="ESTs generated from a normal nasopharynx cDNA
library from southern Chinese"

```

ORIGIN

Query Match	100.0%;	Score 87;	DB 14;	Length 535;
Best Local Similarity	100.0%;	Pred. No. 4.8e-19;		
Matches 87;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY 1 AGAGAACCATATTAAATGAACTGAGATTTTCTGCGCTGAGACTTGACAGGGAGGCAAGA 60  
DB 42 AGAGAACCATATTAAATGAACTGAGATTTTCTGCGCTGAGACTTGACAGGGAGGCAAGA 100

```

QY      61  AGCACTCTGGACACCACTATGGACAG  87
          |||||
Db      102  AGCACTCTGGACACCACTATGGACAG  128

```

RESULT 2	693 bp	linear	EST 15-MAY-2001
6271102F1	693 bp	linear	EST 15-MAY-2001
LOCUS	6271102F1	693 bp	linear
DEFINITION	6271102F1 NTH_MGC_48 Homo sapiens cDNA clone IMAGE:465180 5',		
	mRNA sequence.		

ACCESSION	EG757392	GI:1406045
VERSION	EG757392.1	
KEYWORDS	EST.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 693)  
NIH-MGC <http://mgc.ncl.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished. (1999)  
Contact: Robert Strausberg, Ph.D.

Email: [cqabbs-remail.nih.gov](mailto:cqabbs-remail.nih.gov)  
 Tissue Procurement: Louis W. Staudt, M.D., Ph.D.  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINT)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LINT at:  
<http://image.lnl.gov>  
 Plate: L1CM1694 row: k column: 05  
 High quality sequence stop: 693.  
 Location/Qualifiers

FEATURES

SOURCE

```

1. .693
/organism="Homo sapiens"
/mol_type="rRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4651580"
/tissue_type="Primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"
/clone_1b="NIH MGC 48"
/notes="Organ: B-cells; Vector: pOT87; Site 1: XhoI;
Site 2: EcoRI; cDNA made by oligo-dT priming.
directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GCGCAGAC(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH MGC Library."

```

ORIGIN

Query Match	100.0%;	Score 87;	DB 12;	Length 693;
Best Local Similarity	100.0%;	Pred. No. 5,2e-19;		
Matches 87;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

1 AGAGAACCATTCATTAAATTGAAGGAGATTTTCTGCGCTGAGACTTGCAGGAGGCAGA 60

Db 2 AGAAGACCAATTAATTGTAAGTGAATTTTCTGGCTTGAGCTTGACGGGAGCCAGA 61  
 QY 61 AGACACTTGGACACCACTATGGACG 87  
 Db 62 AGACACTTGGACACCACTATGGACG 88

RESULT	3
BX402063	
LOCUS	
DEFINITION	BX402063 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED EST 13-MAY-2003
	1201 bp mRNA linear
	CDR1 CDR2 CDR3 F RTMRTV RNA sequence

ACCESSION	BX402063
VERSION	BX402063.1
KEYWORDS	GI:30626645
	EST.

ORGANISM	REFERENCE
Human sapiens	1 (Pages 1 to 1201)
Bumakryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.	1 (Pages 1 to 1201)

REFERENCE	1 (bases 1 to 1201)
AUTHORS	Li, W. B., Gruber, C., Jesssee, J. and Polayes, D.
TITLE	Full-length cDNA libraries and normalization
JOURNAL	Unpublished (2001)
COMMENT	Contact: Genoscope
	Genoscope - Centre National de Séquençage

Genoscope, Centre National de Séquençage  
BP 131 006 Evry cedex - France  
Email: [segref@genoscope.cns.fr](mailto:segref@genoscope.cns.fr), Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 6672.r For  
more information about this cluster, see  
[http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0DOL12DB09QPI&cluster=6672.r](http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DOL12DB09QPI&cluster=6672.r). Contact :  
Peng Liang Email : [liang@life tech.com](mailto:liang@life tech.com) URL :  
<http://fulllength.invitrogen.com/> Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0DOL12DB09QPI.

**FEATURES**  
**SOURCE**

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="GSDLO127D18"
/cell_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"
/cell_line="RAMOS CELL LINE"
/clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

```

ORIGIN

Query Match	100.0%;	Score 87;	DB 13;	Length 1201;
Best Local Similarity	100.0%;	Pred. No. 6.2e-19;		
Matches 87;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY 1 AGAAGACATCATATTAATGAATGATTTTCTGCTGAGACTTGAGAGGAGGAGA 60  
Db 73 AGAGACACATCTATTATTAAGTGGATTTTCTGCTGAGACTTGAGAGGAGGAGA 133  
QY 61 AGACACTCTGGACACCATATGCAAG 87  
Db 133 AGACACTCTGGACACCATATGCAAG 159

133 AGACACTCTGGACACCACTATGGACAG 159

## RESULT 4

LOCUS Bg757089 820 bp mRNA linear EST 15-MAY-2001  
DEFINITION 607712124F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4855517 5',  
mRNA sequence.  
ACCESSION Bg757089  
VERSION Bg757089.1 GI:14067742  
KEYWORDS EST.  
SOURCE Homo sapiens (human)



## ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

1 (bases 1 to 820)

NIH-MGC <http://mgs.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)

Tissue Procurement: Louis M. Staudt, M.D., Ph.D.

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

<http://image.llnl.gov>

Plate: LNCM1704 row: 0 column: 06

High quality sequence stop: 675.

Location/Qualifiers

1. 820

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:485517"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: B-cells; Vector: pOTB7; Site\_1: XhoI;

Site\_2: EcoRI; cDNA made by oligo-dT priming.

Directionally cloned into EcoRI/XhoI sites using the

following 5' adaptor: GGCACGAG(G). Size-selected &gt;500bp

for average insert size 1.8kb. Library constructed by Ling

Hong in the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies).

Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 98.2%; Score 85.4; DB 12; Length 820;

Best Local Similarity 98.9%; Pred. No. 2e-18;

Matches 86; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 AGAGAACCATCATTAATTGAAGTATTTCTGCGCTGACACTTGACGAGGACCAAGA 60

6 AGAGAACCATCATTAATTGAAGTATTTCTGCGCTGACACTTGACGAGGACCAAGA 65

61 AGACACTCTGGACACCACTATGACAG 87

66 AGACACTCTGGACACCACTATGACAG 92

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

## FEATURES

source

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

<http://image.llnl.gov>

Plate: LNCM95 row: p column: 20

High quality sequence stop: 541.

Location/Qualifiers

1. 541

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:4054915"

/tissue\_type="primary B-cells from tonsils (cell line)"

/lab\_host="DH10B (phage-resistant)"

/clone\_lib="NIH\_MGC\_48"

/note="Organ: B-cells; Vector: pOTB7; Site\_1: XhoI;

Site\_2: EcoRI; cDNA made by oligo-dT priming.

Directionally cloned into EcoRI/XhoI sites using the

following 5' adaptor: GGCACGAG(G). Size-selected &gt;500bp

for average insert size 1.8kb. Library constructed by Ling

Hong in the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies).

Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 96.6%; Score 84; DB 10; Length 541;

Best Local Similarity 100.0%; Pred. No. 5.2e-18;

Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

4 GAACCATCATTAATTGAAGTATTTCTGCGCTGACACTTGACGAGGACCAAGA 63

2 GAACCATCATTAATTGAAGTATTTCTGCGCTGACACTTGACGAGGACCAAGA 61

64 CACTCTGGACACCACTATGACAG 87

62 CACTCTGGACACCACTATGACAG 85

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db

Db



/clone.lib="NIH\_MGC\_48"  
 /note="Organ: B-cells; Vector: pOTB7; Site 1: XhoI;  
 Site 2: EcoRI; cDNA made by oligo-dT priming.  
 Directionally cloned into EcoRI/XhoI sites using the  
 following 5' adaptor: GGCACGAG(G). Size-selected >500bp  
 for average insert size 1.8kb. Library constructed by Ling  
 Hong in the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies).  
 Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 96.6%; Score 84; DB 12; Length 743;

Best Local Similarity 100.0%; Pred. No. 5.7e-18;  
 Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GAACCATCATTTAATGAAGTGAATTTTCTGGCTGAGACTTGGACGAGGCAAGA 63  
 DB 2 GAACCATCATTTAATGAAGTGAATTTTCTGGCTGAGACTTGGACGAGGCAAGA 61  
 QY 64 CACTCTGACACCACTATGACAG 87  
 DB 62 CACTCTGACACCACTATGACAG 85

## RESULT 7

BP975166 942 bp mRNA linear EST 22-JAN-2001

LOCUS 60224465791 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4335639 5',

DEFINITION mRNA sequence.

ACCESSION BP975166

VERSION BP975166.1 GI:12342381

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE NIH-MGC http://mgc.nci.nih.gov/.

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE Unpublished (1999)

JOURNAL Contact: Robert Strausberg, Ph.D.

COMMENT Email: cgaabs-remail.nih.gov

Tissue Procurement: Louis M. Staudt, M.D., Ph.D.

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCM1207 row: a column: 16

High quality sequence stop: 707.

Location/Qualifiers

1. 942

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:4335639"

/issue\_type="Primary B-cells from tonsils (cell line)"

/lab\_host="DH10B (phage-resistant)"

/clone.lib="NIH\_MGC\_48"

/note="Organ: B-cells; Vector: pOTB7; Site 1: XhoI;  
 Site 2: EcoRI; cDNA made by oligo-dT priming.  
 Directionally cloned into EcoRI/XhoI sites using the  
 following 5' adaptor: GGCACGAG(G). Size-selected >500bp  
 for average insert size 1.8kb. Library constructed by Ling  
 Hong in the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies).  
 Note: this is a NIH\_MGC Library."

Best Local Similarity 100.0%; Pred. No. 6.2e-18;  
 Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## ORIGIN

Query Match 96.6%; Score 84; DB 12; Length 743;

Best Local Similarity 100.0%; Pred. No. 5.7e-18;  
 Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GAACCATCATTTAATGAAGTGAATTTTCTGGCTGAGACTTGGACGAGGCAAGA 63  
 DB 2 GAACCATCATTTAATGAAGTGAATTTTCTGGCTGAGACTTGGACGAGGCAAGA 61  
 QY 64 CACTCTGACACCACTATGACAG 87  
 DB 62 CACTCTGACACCACTATGACAG 85

## RESULT 8

BO065440 953 bp mRNA linear EST 02-APR-2002

LOCUS AGENCOURT 6855061 NIH\_MGC\_99 Homo sapiens cDNA clone IMAGE:5929977

DEFINITION 5', mRNA sequence.

ACCESSION BO065440

VERSION BO065440.1 GI:19894486

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE NIH-MGC http://mgc.nci.nih.gov/.

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE Unpublished (1999)

JOURNAL Contact: Robert Strausberg, Ph.D.

COMMENT Email: cgaabs-remail.nih.gov

Tissue Procurement: Lou Staudt

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCM2108 row: p column: 10

High quality sequence stop: 634.

Location/Qualifiers

1. 953

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:5929977"

/issue\_type="Lymphoma, cell line"

/lab\_host="DH10B (phage-resistant)"

/clone.lib="NIH\_MGC\_99"

/note="Organ: lymph; Vector: pOTB7; Site 1: XhoI; Site 2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
 into EcoRI/XhoI sites using the following 5' adaptor:  
 GGCACGAG(G). Size-selected >500bp for average insert size  
 1.8kb. Library constructed by Ling Hong in the laboratory  
 of Gerald M. Rubin (University of California, Berkeley)  
 using ZAP-cDNA synthesis kit (Stratagene) and Superscript  
 II RT (Life Technologies). Note: this is a NIH\_MGC  
 Library."

Best Local Similarity 100.0%; Pred. No. 1.8e-13;  
 Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## ORIGIN

Query Match 81.6%; Score 71; DB 13; Length 953;

Best Local Similarity 100.0%; Pred. No. 1.8e-13;  
 Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 TTGAAGTGAAGATTTTCTGGCTGAGACTTGGACGAGGCAAGAAGACACTCTGACACC 76  
 DB 1 TTGAAGTGAAGATTTTCTGGCTGAGACTTGGACGAGGCAAGAAGACACTCTGACACC 60  
 QY 77 ACTATGACAG 87  
 DB 61 ACTATGACAG 71

## RESULT 9

Query Match 81.6%; Score 71; DB 13; Length 953;

Best Local Similarity 100.0%; Pred. No. 1.8e-13;  
 Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match 96.6%; Score 84; DB 10; Length 942;



B0055935  
 LOCUS B0055935 1052 bp mRNA linear EST 29-MAR-2002  
 DEFINITION AGNCOURT 6796291 NIH\_MGC\_99 Homo sapiens cDNA clone IMAGE:5808181  
 5', mRNA sequence.  
 ACCESSION B0055935  
 VERSION B0055935.1 GI:19815262  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 1052)  
 NIH-MGC <http://mgi.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: Lou Staudt  
 cDNA Library Preparation: Rubin Laboratory  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: L1CM2051 row: m column: 14  
 High quality sequence stop: 665.  
 Location/Qualifiers  
 1..1052  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5808181"  
 /tissue\_type="lymphoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_1lb="NIH\_MGC\_99"  
 /note="Organ: lymph; Vector: pOTB7, Site 1: XhoI; Site 2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
 into EcoRI/XhoI sites using the following 5' adaptor:  
 GGCACGAG(G). Size-selected >500bp for average insert size  
 1.8kb. Library constructed by Ling Hong in the laboratory  
 of Gerald M. Rubin (University of California, Berkeley)  
 using ZAP-cDNA synthesis kit (Stratagene) and Superscript  
 II RT (Life Technologies). Note: this is a NIH\_MGC  
 Library."

ORIGIN  
 Query Match 81.6%; Score 71; DB 13; Length 1052;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-13;  
 Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 TTGAAGTGAATTTTTCGGCCTGAGACTTCAGAGAGGCAAGACACTCTGGAAC 76  
 Db 1 TTGAAGTGAATTTTTCGGCCTGAGACTTCAGAGAGGCAAGACACTCTGGAAC 60

QY 77 ACTATGACAG 87  
 Db 61 ACTATGACAG 71

RESULT 10  
 LOCUS B0758510 872 bp mRNA linear EST 15-MAY-2001  
 DEFINITION 60271272121 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4853069 5',  
 mRNA sequence.  
 ACCESSION B0758510  
 VERSION B0758510.1 GI:14069163  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 872)  
 NIH-MGC <http://mgi.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: L1CM1638 row: i column: 06  
 High quality sequence stop: 836.  
 Location/Qualifiers  
 1..872  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4853069"  
 /tissue\_type="Primary B-cells from tonsils (cell line)"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_1lb="NIH\_MGC\_48"  
 /note="Organ: B-cells; Vector: pOTB7, Site 1: XhoI;  
 Site 2: EcoRI; cDNA made by oligo-dT priming.  
 Directionally cloned into EcoRI/XhoI sites using the  
 following 5' adaptor: GGCACGAG(G). Size-selected >500bp  
 for average insert size 1.8kb. Library constructed by Ling  
 Hong in the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies).  
 Note: this is a NIH\_MGC Library."

ORIGIN  
 Query Match 80.9%; Score 70.4; DB 12; Length 872;  
 Best Local Similarity 98.6%; Pred. No. 2.8e-13;  
 Matches 71; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 16 ATTGAAGTGAATTTTTCGGCCTGAGACTTCAGAGAGGCAAGACACTCTGGAAC 75  
 Db 2 ATTGAAGTGAATTTTTCGGCCTGAGACTTCAGAGAGGCAAGACACTCTGGAAC 61

QY 76 CACTATGACAG 87  
 Db 62 CACTATGACAG 73

RESULT 11  
 LOCUS AL559877 1201 bp mRNA linear EST 31-MAY-2003  
 DEFINITION AL559877 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA  
 clone CG00D003YB14 5-PRIME, mRNA sequence.  
 ACCESSION AL559877  
 VERSION AL559877.2 GI:31284008  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 1201)  
 Li W.B., Gruber C., Jesse J. and Polayes D.  
 Full-length cDNA libraries and normalization  
 Unpublished (2001)  
 On Feb 15, 2001 this sequence version replaced gi:12905793.  
 Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: [secref@genoscope.cns.fr](mailto:secref@genoscope.cns.fr), Web: [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 6672.r For  
 more information about this cluster, see  
<http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CG00D003YB14>  
 Feng Liang Email: [fliang@lifeotech.com](mailto:fliang@lifeotech.com) URL:



http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0D0003A07Qp1.

## FEATURES

Location/Qualifiers  
1..1201

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="CS0D0003YB14"

/tissue\_type="B CELLS (RAMOS CELL LINE)"

/cell\_line="RAMOS CELL LINE"

/clone\_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"

/note="Vector: pCMVSPORT 6, 1st strand cDNA was primed

with a NotI-oligo (dT) primer. Five prime end enriched,

double-strand cDNA was digested with Not I and cloned into

the Not I and EcoRV sites of the pCMVSPORT 6 vector.

Library was not normalized."

## ORIGIN

Query Match 45.1%; Score 39.2; DB 9; Length 1201;

Best Local Similarity 73.5%; Pred. No. 0.015; Indels 0; Gaps 0;

Matches 50; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

20 AAGTGAATTTTCTGGCCTTGAAGACTTGACAGGAGGAGAGACACTTGACACCACT 79

32 AAGCAGCTGTGACCGGTCCGGAATTCCTCGGATGGAGAGACACTTGACACCACT 91

80 ATGCACAG 87

92 ATGCACAG 99

## RESULT 12

BX464579 853 bp mRNA linear EST 22-MAY-2003

LOCUS BX464579 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA

DEFINITION clone CS0D0003YB14 5-PRIME, mRNA sequence.

ACCESSION BX464579

VERSION BX464579.1 GI:31031641

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 853)

Li, W.-B., Gruber, C., Jessee, J., and Polayes, D.

Full-length cDNA libraries and normalization

Unpublished (2001)

Contact: Genoscope

Genoscope - Centre National de Sequencage

BP 191 91006 Evry cedex - France

Email: seqref@genoscope.cns.fr; Web : www.genoscope.cns.fr

Library was constructed by Life Technologies, a division of

Invitrogen. This sequence belongs to sequence cluster 6672.r For

more information about this cluster, see

http://www.genoscope.cns.fr/

cgi-bin/cluster.cgi?seq=CS0D0003YB14&cluster=6672.r. Contact :

Peng Liang Email : filiang@life.techn.com URL :

http://fulllength.invitrogen.com/ Invitrogen Corporation 1600

Faraday Avenue Genoscope sequence ID : CS0D0003YB14Qp1.

Location/Qualifiers

1..853

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="CS0D0003YB14"

/tissue\_type="B CELLS (RAMOS CELL LINE)"

/cell\_line="RAMOS CELL LINE"

/clone\_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"

/note="Vector: pCMVSPORT 6, 1st strand cDNA was primed

with a NotI-oligo (dT) primer. Five prime end enriched,

double-strand cDNA was digested with Not I and cloned into

the Not I and EcoRV sites of the pCMVSPORT 6 vector.

Library was not normalized."

## ORIGIN

Query Match 44.6%; Score 38.8; DB 13; Length 853;

Best Local Similarity 72.1%; Pred. No. 0.019; Indels 0; Gaps 0;

Matches 49; Conservative 1; Mismatches 18; Indels 0; Gaps 0;

20 AAGTGAATTTTCTGGCCTTGAAGACTTGACAGGAGGAGAGACACTTGACACCACT 79

27 AAGCAGCTGTGACCGGTCCGGAATTCCTCGGATGGAGAGACACTTGACACCACT 86

80 ATGCACAG 87

87 ATGCACAG 94

## RESULT 13

BG686876 889 bp mRNA linear EST 01-MAY-2001

LOCUS BG686876 602650861P1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4763247 5'

DEFINITION mRNA sequence.

ACCESSION BG686876

VERSION BG686876.1 GI:13918273

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 889)

NIH-MGC http://mgi.nci.nih.gov/

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgaabs@mail.nih.gov

Tissue Procurement: Louis M. Staudt, M.D., Ph.D.

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

http://image.llnl.gov

Plate: LNCM1618 row: 1 column: 16

High quality sequence start: 6

High quality sequence stop: 727.

Location/Qualifiers

1..889

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:4763247"

/tissue\_type="Primary B-cells from tonsils (cell line)"

/lab\_host="PH10B (phage-resistant)"

/clone\_lib="NIH\_MGC\_48"

/note="Organ: B-cells; Vector: pOTB7, Site:1; XhoI;

Site:2; EcoRI; cDNA made by oligo-dT priming.

Directionally cloned into EcoRI/XhoI sites using the

following 5' adaptor: GGCACGAG(G). Size-selected >500bp

for average insert size 1.8kb. Library constructed by Ling

Hong in the laboratory of Gerald W. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies).

Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 42.5%; Score 37; DB 12; Length 889;

Best Local Similarity 100.0%; Pred. No. 0.08; Indels 0; Gaps 0;

Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

51 GGAGGAGAGAGACACTTGACACCACTATGACAG 87

9 GGAGGAGAGAGACACTTGACACCACTATGACAG 45

## RESULT 14



CC198363/c 930 bp DNA linear GSS 22-MAY-2003  
 LOCUS ZM98363.1 ZM98363.1 ZM98363.1 ZM98363.1  
 DEFINITION ZM98363.1 ZM98363.1 ZM98363.1 ZM98363.1  
 ACCESION CC198363.1 CC198363.1 CC198363.1 CC198363.1  
 VERSION CC198363.1 CC198363.1 CC198363.1 CC198363.1  
 KEYWORDS GSS.  
 SOURCE Zea mays subsp. mays (maize)  
 ORGANISM Zea mays subsp. mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoidae; Andropogoneae; Zea.  
 1 (bases 1 to 930)  
 Bhatti,A.K., Young,S., Kavchok,S., Keizer,G., Bronzino,A.C.,  
 Sudhiana,R., Fuke,G., Yu,Y., Wang,R. and Messing,J.  
 Sequencing of the maize genome at PCR (2003a)  
 Unpublished (2003)  
 CONTACT Bhatti,A.K.  
 Dr. Joachim Messing's lab  
 The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers  
 University  
 190 Frelinghuysen Road, Piscataway, NJ 08854, USA  
 Tel: 732 445 3801  
 Fax: 732 445 5735  
 Email: bhatti@waksman.rutgers.edu  
 Seq primer: SP6  
 Class: BAC ends  
 High quality sequence start: 3.  
 Location/Qualifiers  
 1..930  
 /organism="Zea mays subsp. mays"  
 /mol\_type="genomic DNA"  
 /cultivar="B73"  
 /sub\_species="mays"  
 /db\_xref="taxon:4578"  
 /clone="ZM98363.1L03"  
 /lab\_host="E. coli DH10B"  
 /clone\_id="ZM98363"  
 /note="Vector: pTAKBAC1.3; Site\_1: BamHI; Site\_2: BamHI"

ORIGIN  
 Query Match 36.3%; Score 31.6; DB 28; Length 930;  
 Best Local Similarity 65.7%; Pred. No. 5.8; Mismatches 24; Indels 0; Gaps 0;  
 Matches 46; Conservative 0; Mismatches 24; Indels 0; Gaps 0;  
 QY 1 AGAGAACCATCAATTAATGAGTGAATTTCTGCTGAGACTTGACGAGGAGGACAGA 60  
 DB 763 AGAGAACCATCAATTAATGAGTGAATTTCTGCTGAGACTTGACGAGGAGGACAGA 724  
 QY 61 AGACACTCTG 70  
 DB 723 AGCCACTGTG 714

RESULT 15  
 BE071397 250 bp mRNA linear EST 09-JUN-2000  
 LOCUS RC2-BT0506-240400-016-g05 BT0506 Homo sapiens CDNA, mRNA sequence.  
 DEFINITION RC2-BT0506-240400-016-g05 BT0506 Homo sapiens CDNA, mRNA sequence.  
 ACCESION BE071397  
 VERSION BE071397.1 GI:8416445  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 250)  
 Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
 Nagai,M.A., da Silva,M. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,  
 Brunstein,A., de Oliveira,P.S., Bucher,P., Jongeneel,C.V.,  
 O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
 Simpson,A.J.  
 Shotgun sequencing of the human transcriptome with ORF expressed

JOURNAL sequence tags  
 MEDLINE Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 PUBMED 20202663  
 10737800  
 CONTACT Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 Project. This entry can be seen in the following URL  
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?cl=RC2-BT0506-240400-016-g05&cl2=2000-04-24&cl4=1>)  
 Seq primer: puc 18 forward  
 High quality sequence start: 26  
 High quality sequence stop: 250.  
 Location/Qualifiers  
 1..250  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /dev\_stage="Adult"  
 /clone\_id="BT0506"  
 /note="Organ: breast; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN  
 Query Match 35.4%; Score 30.8; DB 10; Length 250;  
 Best Local Similarity 63.5%; Pred. No. 7.2; Mismatches 27; Indels 0; Gaps 0;  
 Matches 47; Conservative 0; Mismatches 27; Indels 0; Gaps 0;  
 QY 4 GAACCATCAATTAATGAGTGAATTTCTGCTGAGACTTGACGAGGAGGACAGA 63  
 DB 123 GAACCATCAATTAATGAGTGAATTTCTGCTGAGACTTGACGAGGAGGACAGA 64  
 QY 64 CACTTGACACCA 77  
 DB 63 CAGCTGAACATCA 50

Search completed: March 13, 2004, 00:38:59  
 Job time : 280.155 secs







DB 1 AGAGACCATCATTTAATTGAAAGTGAAGATTTTCTGGCCTGAGACTTGAGGAGGCAAGA 60  
QY 61 AGACACTCTGGACACCACTATGAGACAG 87  
DB 61 AGACACTCTGGACACCACTATGAGACAG 87

RESULT 2  
BD016833  
LOCUS Novel cytidine deaminase. 2818 bp DNA linear PAT 27-AUG-2002  
DEFINITION BD016833  
ACCESSION BD016833.1 GI:22558009  
VERSION JP 2001245669-A/6.  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS Honjo, T. and Muramatsu, M.  
TITLE Novel cytidine deaminase  
JOURNAL Patent: JP 2001245669-A 6 11-SEP-2001;  
JAPAN TOBACCO INC, TASUKU HONJO  
OS Homo sapiens (human)  
PN JP 2001245669-A/6  
PD 11-SEP-2001  
PF 28-MAR-2000 JP 2000092981  
PI TASUKU HONJO, MASAMICHI MURAMATSU  
PC C12N15/09, A61K39/395, A61K39/395, A61P1/00, A61P11/06, A61P13/12,  
A61P17/00, A61P27/02, A61P37/08, C07K16/18, C12N1/19, PC  
C12N1/21,  
PC C12N5/10, C12N9/78, C12P21/02, C12P21/08, C12N1/21, C12R1/19, PC  
(C12N5/10, C12R1/91), C12N15/00, C12N5/00, (C12N5/00, C12R1/91) CC  
FH Key Location/Qualifiers  
FT 5'UTR (1), (79)  
FT CDS (80), (676)  
FT 3'UTR (677), (2818).  
FEATURES  
source Location/Qualifiers  
1..2818  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

ORIGIN  
Query Match 100.0%; Score 87; DB 6; Length 2818;  
Best Local Similarity 100.0%; Pred. No. 8,7e-41;  
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAGACCATCATTTAATTGAAAGTGAAGATTTTCTGGCCTGAGACTTGAGGAGGCAAGA 60  
DB 1 AGAGACCATCATTTAATTGAAAGTGAAGATTTTCTGGCCTGAGACTTGAGGAGGCAAGA 60

QY 61 AGACACTCTGGACACCACTATGAGACAG 87  
DB 61 AGACACTCTGGACACCACTATGAGACAG 87

RESULT 3  
BD016834  
LOCUS Novel cytidine deaminase. 5514 bp DNA linear PAT 27-AUG-2002  
DEFINITION BD016834  
ACCESSION BD016834.1 GI:22558010  
VERSION JP 2001245669-A/7.  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS Honjo, T. and Muramatsu, M.  
TITLE Novel cytidine deaminase  
JOURNAL Patent: JP 2001245669-A 7 11-SEP-2001;

COMMENT JAPAN TOBACCO INC, TASUKU HONJO  
OS Homo sapiens (human)  
PN JP 2001245669-A/7  
PD 11-SEP-2001  
PF 28-MAR-2000 JP 2000092981  
PI TASUKU HONJO, MASAMICHI MURAMATSU  
PC C12N15/09, A61K39/395, A61K39/395, A61P1/00, A61P11/06, A61P13/12,  
A61P17/00, A61P27/02, A61P37/08, C07K16/18, C12N1/19, PC  
C12N1/21,  
PC C12N5/10, C12N9/78, C12P21/02, C12P21/08, C12N1/21, C12R1/19, PC  
(C12N5/10, C12R1/91), C12N15/00, C12N5/00, (C12N5/00, C12R1/91) CC  
FH Key Location/Qualifiers  
FT 5'UTR (1), (1031)  
FT exon (1032), (1118)  
FT intron (1119), (5514).  
FEATURES  
source Location/Qualifiers  
1..5514  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

ORIGIN  
Query Match 100.0%; Score 87; DB 6; Length 5514;  
Best Local Similarity 100.0%; Pred. No. 8e-41;  
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAGACCATCATTTAATTGAAAGTGAAGATTTTCTGGCCTGAGACTTGAGGAGGCAAGA 60  
DB 1032 AGAGACCATCATTTAATTGAAAGTGAAGATTTTCTGGCCTGAGACTTGAGGAGGCAAGA 1091

QY 61 AGACACTCTGGACACCACTATGAGACAG 87  
DB 1092 AGACACTCTGGACACCACTATGAGACAG 1118

RESULT 4  
BD016860  
LOCUS Novel cytidine deaminase. 11204 bp DNA linear PAT 27-AUG-2002  
DEFINITION BD016860  
ACCESSION BD016860.1 GI:22558036  
VERSION JP 2001245669-A/33.  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS Honjo, T. and Muramatsu, M.  
TITLE Novel cytidine deaminase  
JOURNAL Patent: JP 2001245669-A 33 11-SEP-2001;  
JAPAN TOBACCO INC, TASUKU HONJO  
OS Homo sapiens (human)  
PN JP 2001245669-A/33  
PD 11-SEP-2001  
PF 28-MAR-2000 JP 2000092981  
PI TASUKU HONJO, MASAMICHI MURAMATSU  
PC C12N15/09, A61K39/395, A61K39/395, A61P1/00, A61P11/06, A61P13/12,  
A61P17/00, A61P27/02, A61P37/08, C07K16/18, C12N1/19, PC  
C12N1/21,  
PC C12N5/10, C12N9/78, C12P21/02, C12P21/08, C12N1/21, C12R1/19, PC  
(C12N5/10, C12R1/91), C12N15/00, C12N5/00, (C12N5/00, C12R1/91) CC  
FH Key Location/Qualifiers  
1..11204  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

ORIGIN  
Query Match 100.0%; Score 87; DB 6; Length 11204;  
Best Local Similarity 100.0%; Pred. No. 7.4e-41;



Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAGAACCATCATTAATTGAGATTTTCTGGCTGAGACTTGACGAGGAGCAAGA 60  
DB 442 AGAGAACCATCATTAATTGAGATTTTCTGGCTGAGACTTGACGAGGAGCAAGA 501

QY 61 AGACACTCTGGACACCATATGACAG 87  
DB 502 AGACACTCTGGACACCATATGACAG 528

RESULT 5  
AB040430 11204 bp DNA linear PRI 03-OCT-2000  
AB040430  
LOCUS Homo sapiens AID gene for activation-induced cytidine deaminase,  
DEFINITION complete cds.  
ACCESSION AB040430  
VERSION AB040430.1 GI:9988407  
KEYWORDS AID; activation-induced cytidine deaminase.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Muto,T., Muramatsu,M., Taniwaki,M., Kinoshita,K. and Honjo,T.  
TITLE Isolation, tissue distribution, and chromosomal localization of the  
human activation-induced cytidine deaminase (AID) gene  
JOURNAL Genomics 68 (1), 85-88 (2000)  
MEDLINE 20408890  
PUBMED 10950930  
REFERENCE  
AUTHORS Revy,P., Muto,T., Levy,Y., Gelsmann,F., Plehani,A., Sanal,O.,  
Cattan,N., Forveille,M., Dufourcq-Iagelouse,R., Gernery,A.,  
Teezan,I., Ereyo,F., Kayseril,H., Ugazio,A.G., Brousse,N.,  
Muramatsu,M., Notarangelo,L.D., Kinoshita,K., Honjo,T., Fischer,A.  
and Durandy,A.  
TITLE Activation-induced cytidine deaminase (AID) deficiency causes the  
autosomal recessive form of the Hyper-IgM syndrome (HIGM2)  
JOURNAL Cell 102 (5), 565-575 (2000)  
MEDLINE 20460541  
PUBMED 11007475  
REFERENCE  
AUTHORS Muto,T., Muramatsu,M., Taniwaki,M., Kinoshita,K. and Honjo,T.  
TITLE Direct Submission  
Submitted (18-MAR-2000) Tasuku Honjo, Kyoto University, Department  
of Medical Chemistry, Faculty of Medicine, Yoshida, Sakyo-ku,  
Kyoto, Kyoto 606-8501, Japan (E-mail:honjo@four.med.kyoto-u.ac.jp,  
Tel:81-75-753-4371(ex.4371), Fax:81-75-753-4388)  
FEATURES  
source location/Qualifiers  
1..11204  
/organism="Homo sapiens"  
/mol\_type="Genomic DNA"  
/db\_xref="taxon:9606"  
join(521..528,6280..6427,7807..8077,8371..8486,8956..9009)  
/gene="AID"  
join(521..528,6280..6427,7807..8077,8371..8486,8956..9009)  
/gene="AID"  
/codon\_start=1  
/product="activation-induced cytidine deaminase"  
/protein\_id="BAB12720.1"  
/db\_xref="GI:9988408"  
/translation="MDSLMMNRKFLYQFNKVRKGRRETYLYCYVVRKDSATSFSL  
DFGYLRNKGCHVEFLFRLYSMDLDPGRCYRVWTSNPGPCARHAFVLRGMP  
NLRIPTARLYFCEDRKAEPEGRLRLRAGVQALWTFKQYFCVMTFVENHRTFK  
AMEGHHSVLSRLRRLRLPLIVVDLRARFTLIG"

ORIGIN  
Query Match 100.0%; Score 87; DB 9; Length 11204;  
Best Local Similarity 100.0%; Pred. No. 7,4e-41;  
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 AGAGAACCATCATTAATTGAGATTTTCTGGCTGAGACTTGACGAGGAGCAAGA 60  
|||||  
442 AGAGAACCATCATTAATTGAGATTTTCTGGCTGAGACTTGACGAGGAGCAAGA 60  
|||||

DB 442 AGAGAACCATCATTAATTGAGATTTTCTGGCTGAGACTTGACGAGGAGCAAGA 501

QY 61 AGACACTCTGGACACCATATGACAG 87  
DB 502 AGACACTCTGGACACCATATGACAG 528

RESULT 6  
AC092184 71132 bp DNA linear PRI 12-JUN-2002  
AC092184  
LOCUS Homo sapiens 12 BAC RP11-43817 (Roswell Park Cancer Institute Human  
DEFINITION BAC library) complete sequence.  
ACCESSION AC092184  
VERSION AC092184.7 GI:21206067  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Al-Osman,F.R., Allen,C.,  
Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T.,  
Barbieri,J., Benton,J., Bimaye,K., Blankenburg,K., Bonnin,D.,  
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,  
Buhay,C., Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
Chen,G., Chen,R., Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C.,  
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Dublin,K.J.,  
Eberhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Emerling,S.,  
Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P.,  
Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N.,  
Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S.,  
Hamilton,K., Han,J., Harris,K., Harris,K., Hart,M., Havlak,F.,  
Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M.,  
Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,U., Huliy,S.,  
Hume,J., Ioshikhes,I., Jackson,L.E., Jacobson,B., Jia,Y.,  
Johnson,R., Jollivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U.,  
King,L., Korvav,J., Kovar,C., Kratovic,J., Kureishi,A., Landry,N.,  
Leal,B., Lee,E., Lewis,L.C., Lewis,L., Li,J., Li,Z., Licharge,O.,  
Lieu,C., Liu,J., Liu,W., Loussed,H., Lozano,R.J., Lu,X.,  
Lucier,A., Lucier,R., Luna,R., Ma,J., Maneswari,M., Mapa,P.,  
Marondei,I., Martin,R., Martindale,A., Martinez,E., Massey,E.,  
Mawker,M., Miller,A., Miner,G., Miner,Z., Mitchell,T.,  
Mozhabat,K., Montgomery,K.T., Morgan,M., Morris,S., Moser,M.,  
Neal,D., Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,  
Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogih,M., Okunou,G.,  
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
Peters,L., Pickens,R., Primus,B., Pu,L.L., Quiles,M., Ren,Y.,  
Rivers,M., Rojas,A., Rojiboken,I., Rolfe,M., Ruiz,S., Savery,G.,  
Scheer,S., Scott,G., Shen,H., Shim,C., Shooharri,N., Sisson,I.,  
Sodergren,E., Sonaike,I., Sparks,A., Stanley,H., Stone,H.,  
Sutton,A., Svatek,A., Tabori,P., Tamerisa,A., Tamerisa,K., Tang,H.,  
Tansley,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,  
Uemami,K., Vasquez,L., Vera,V., Vallation,D., Vinson,R., Wang,Q.,  
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlingon,S.,  
Williams,G., Williamson,A., Wleczek,R., Wooten,S., Worley,K.,  
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorilla,S., Kucherlapati,R.,  
Weinstock,G. and Gibbs,R.  
TITLE Direct Submission  
JOURNAL Unpublished  
AUTHORS Worley,K.C.  
REFERENCE Direct Submission  
TITLE Direct Submission  
Submitted (25-JUN-2001) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 71132)  
Worley,K.C.  
TITLE Direct Submission



JOURNAL Submitted (18-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 4 (bases 1 to 71132)

AUTHORS Morley,K.C.

TITLE Direct Submission

JOURNAL Submitted (25-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 5 (bases 1 to 71132)

AUTHORS Morley,K.C.

TITLE Direct Submission

JOURNAL Submitted (12-JUN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT On May 25, 2002 this sequence version replaced gi:20901754. INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email [gc-help@bcm.tmc.edu](mailto:gc-help@bcm.tmc.edu)

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

#### ANNOTATION OF FEATURES:

STSs are identified using ePCR (genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

#### QUALSTAT-REPORT.

FEATURES

source Location/Qualifiers

1..71132

    /organism="Homo sapiens"

    /mol\_type="genomic DNA"

    /db\_xref="taxon:9606"

    /chromosome="12"

    /clone="RP11-438L7"

    /complement(1..1448)

    /notes="overlaps bases 1..1448 of clone AC092490"

    /function="Clone overlap"

    misc\_feature

        STS

            /standard\_name="57233"

            439..560

        STS

            /standard\_name="92005"

            774..881

        repeat\_region

            /rpt\_family="MIR"

            903..1190

        repeat\_region

            /rpt\_family="Alusq"

            1191..1213

STS /rpt\_family="AT\_rich"

STS 1591..1807

STS /standard\_name="6612"

STS 1744..1819

STS /standard\_name="6198"

repeat\_region 1966..2264

repeat\_region /rpt\_family="Alusq"

repeat\_region 3296..3328

repeat\_region /rpt\_family="(TTTC)n"

repeat\_region 4097..4249

repeat\_region /rpt\_family="Alusq"

repeat\_region 4360..4654

repeat\_region /rpt\_family="AluY"

repeat\_region complement(5140..5262)

repeat\_region /rpt\_family="FLAM\_C"

repeat\_region 5275..5304

repeat\_region /rpt\_family="AT\_rich"

repeat\_region 5305..5563

repeat\_region /rpt\_family="AluA"

repeat\_region 5564..5599

repeat\_region /rpt\_family="AT\_rich"

repeat\_region 5644..5733

repeat\_region /rpt\_family="MSTD"

repeat\_region 5734..6012

repeat\_region /rpt\_family="Alusq"

repeat\_region 6020..6336

repeat\_region /rpt\_family="AlusX"

repeat\_region 6337..6629

repeat\_region /rpt\_family="MSTD"

repeat\_region 6630..6650

repeat\_region /rpt\_family="(CNA)n"

repeat\_region complement(6711..6817)

repeat\_region /rpt\_family="XIR"

repeat\_region complement(7033..7327)

repeat\_region /rpt\_family="AlusX"

repeat\_region 7607..7745

repeat\_region /rpt\_family="FLAM\_A"

repeat\_region 7980..8047

repeat\_region /rpt\_family="MIR"

repeat\_region complement(8050..8549)

repeat\_region /rpt\_family="LTR47A"

repeat\_region 8567..8698

repeat\_region /rpt\_family="MIR"

repeat\_region complement(8815..9101)

repeat\_region /rpt\_family="Alusq"

repeat\_region complement(11227..12175)

repeat\_region /rpt\_family="LTR5"

STS 12625..12900

STS /standard\_name="87432"

repeat\_region complement(13547..13664)

repeat\_region /rpt\_family="LIMB5"

repeat\_region complement(13665..13979)

repeat\_region /rpt\_family="AluY"

repeat\_region complement(13980..14193)

repeat\_region /rpt\_family="LIMB5"

repeat\_region 14622..14924

repeat\_region /rpt\_family="AluY"

repeat\_region complement(15813..17136)

repeat\_region /rpt\_family="LIPAB"

repeat\_region complement(17272..17395)

repeat\_region /rpt\_family="FLAM\_C"

repeat\_region 17426..17575

repeat\_region /rpt\_family="MIR"

Query Match 100.0%; Score 87; DB 9; Length 71132;

Best Local Similarity 100.0%; Pred. No. 5,9e-41;

Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAGAACATCATTAATGAGAGAGATTTTCTGCGCCGAGACTTGCGAGGAGGCAAGA 60

DB 34990 AGAGAACATCATTAATGAGAGATTTTCTGCGCCGAGACTTGCGAGGAGGCAAGA 35049

QY 61 AGACACTCTGACACCACTATGACAG 87



DB 35050 AGACACTGTGACACCATATGACAG 35076

RESULT 7  
LOCUS BC006296  
DEFINITION Homo sapiens activation-induced cytidine deaminase, mRNA (CDNA)  
ACCESSION BC006296  
VERSION BC006296.2 GI:33871601  
KEYWORDS MGC.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1828)  
Straussberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Heide, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hult, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butler, Y.S., Schein, J.E., Jones, S.J., and Marra, M.A.  
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

TITLE  
JOURNAL MEDLINE  
PUBMED 22388257  
2 (bases 1 to 1828)  
Straussberg, R.  
Direct Submission  
Submitted (09-APR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK  
COMMENT NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
On Aug 19, 2003 this sequence version replaced gi:13623400.  
Contact: MGC help desk  
Email: [cgaps-rc@mail.nih.gov](mailto:cgaps-rc@mail.nih.gov)  
Tissue Procurement: Louis Staudt  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),  
Gaithersburg, Maryland;  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc\\_mgc@hgrl.nih.gov](mailto:nisc_mgc@hgrl.nih.gov)  
Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Binkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Lari, P., Legaspi, R., Maduro, O.L., Masello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantipop, S., Thomas, P.J., Touchman, J.W., Tsuneyon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>  
Series: IRAL Plate: 17 Row: a Column: 1  
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 10190699.

FEATURES  
SOURCE  
Location/Qualifiers  
1..1828  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="MGC:12911 IMAGE:4054915"  
/issue\_type="Primary B-Cells from Tonsils"  
/clone\_id="NIH\_MGC\_48"  
/lab\_host="DH10B-R"  
/note="Vector: POT87"  
1..1828  
/gene="AIDCA"  
/note="Synonyms: AID, HIGM2, CDA2, ARP2"  
/db\_xref="LocusID:57379"  
/db\_xref="MIM:605257"  
77..673  
/codon\_start=1  
/product="activation-induced cytidine deaminase"  
/protein\_id="AAH06296.1"  
/db\_xref="GI:13623401"  
/db\_xref="LocusID:57379"  
/translation="MDSILNMRKFLPYOFANVAKGRETYLYVKRDSATSL  
DFGVLKRNKCHLELFLRYSIDMDIDPRCYVMTSPCYDCAHVDRLGDP  
NLSRLFTALVIFCEDRKAPBGLRLHRAGVQIAIMTRKDYFCMTVEVNERIFK  
AMEGLHNSVRLSRQLRLILPLVEVDLDRAFTLGL"

ORIGIN  
Query Match 96.6%; Score 84; DB 9; Length 1828;  
Best Local Similarity 100.0%; Pred.No. 5.7e-39; Indels 0; Gaps 0;  
Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
4 GAACCATCATTAATTGAGTGAATTTTCTGGCTGTGAGACTTGACGGAGGACAGAGA 63  
1 GAACCATCATTAATTGAGTGAATTTTCTGGCTGTGAGACTTGACGGAGGACAGAGA 60

DB 64 CACTCTGACACCATATGACAG 87  
61 CACTCTGACACCATATGACAG 84

RESULT 8  
LOCUS AB040431  
DEFINITION Homo sapiens AID mRNA for activation-induced cytidine deaminase, complete cds.  
ACCESSION AB040431  
VERSION AB040431.1 GI:9988409  
KEYWORDS AID; activation-induced cytidine deaminase; Human AID.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1828)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 1828)  
Muro, T., Muramatsu, M., Taniwaki, M., Kinoshita, K. and Honjo, T.  
Isolation, tissue distribution, and chromosomal localization of the human activation-induced cytidine deaminase (AID) gene  
Genomics 68 (1), 85-88 (2000)

JOURNAL MEDLINE  
PUBMED 20408690  
2 (bases 1 to 1828)  
Remy, P., Muro, T., Levy, Y., Geissmann, F., Plebani, A., Sanal, O., Catalan, N., Forveille, M., Dufourcq-Lageat, R., Gennery, A., Tescan, I., Ersoy, F., Kayserili, H., Ugazio, A.G., Brousse, N., Tezcan, I., Muro, T., Notarangelo, L.D., Kinoshita, K., Honjo, T., Fischer, A. and Durandy, A.  
Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IGM syndrome (HIGM2)  
Cell 102 (5), 565-575 (2000)

JOURNAL MEDLINE  
PUBMED 20460541  
3 (bases 1 to 2791)  
Muro, T., Muramatsu, M., Taniwaki, M., Kinoshita, K. and Honjo, T.  
Direct Submission



## JOURNAL

Submitted (18-MAR-2000) Tasuku Honjo, Kyoto University, Department of Medical Chemistry, Faculty of Medicine, Yoshida, Sakyo-ku, Kyoto 606-8501, Japan (E-mail: honjo@cour.med.kyoto-u.ac.jp, Tel:81-75-753-4371 (ex.4371), Fax:81-75-753-4388)

## FEATURES

## source

1. .2791  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
1. .2791  
/gene="AID"  
77. 673  
/gene="AID"  
/codon\_start=1  
/product="activation-induced cytidine deaminase"  
/protein\_id="BAB12721.1"  
/db\_xref="GI:9988410"  
/translation="MDSILMRKPIYQKQVWMAKGRRETYLGVVRRDSATSFSL  
DQYLRNKGCHVELFLRYISDMDIDPRCVRVTFMSPCDCAHVAADPLRGNP  
NLSRLFTALYFCEDRKAPGGLRLHRAVGQIALMTFKDYFCMTFVNEHETFK  
AMEGLHENSVRISRLRLRLILPLVEYDILRLDAFRLGL"

## CDS

## ORIGIN

Query Match 96.6%; Score 84; DB 9; Length 2791;  
Best Local Similarity 100.0%; Pred. No. 5,4e-39;  
Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 GAACATCATTAATGAGAGATTTTCTGCGCTGAGACCTGACGAGGACAGAGA 63

Db 1 GAACATCATTAATGAGAGATTTTCTGCGCTGAGACCTGACGAGGACAGAGA 60

Qy 64 CACTCTGACACCACTATGACAG 87

Db 61 CACTCTGACACCACTATGACAG 84

RESULT 9 BD016856 26 bp DNA linear PAT 27-AUG-2002

## LOCUS

BD016856 Novel cytidine deaminase. 26 bp

BD016856 BD016856.1 GI:22558032

VERSION JP 2001245669-A/29.

KEYWORDS synthetic construct

SOURCE synthetic construct

ORGANISM artificial sequences.

REFERENCE 1 (bases 1 to 26)

AUTHORS Honjo, T. and Muramatsu, M.

TITLE Novel cytidine deaminase

JOURNAL Patent: JP 2001245669-A 29 11-SEP-2001;

COMMENT JAPAN TOBACCO INC, TASUKU HONJO

OS Artificial Sequence

PN JP 2001245669-A/29

PD 11-SEP-2001

PF 28-MAR-2000 JP 2000092991

PI TASUKU HONJO, MASAMICHI MURAMATSU

PC C12N15/09, A61K39/395, A61K39/395, A61P1/00, A61P11/06, A61P13/12,

PC A61P17/00,

PC A61P27/02, A61P27/16, A61P37/02, A61P37/08, C07K16/18, C12N1/19, PC

C12N1/21,

PC C12N5/10, C12N9/78, C12P21/02, C12P21/08, C12N1/21, C12R1/19, PC

(C12N5/10, C12R1/91), C12N15/00, C12N5/00, C12N5/00, C12R1/91) CC

Description of Artificial Sequence: Artificially synthesized CC

primer

CC sequence, p48

PH Key Location/Qualifiers.

FEATURES

source

1. .26 Location/Qualifiers

/organism="synthetic construct"

/mol\_type="genomic DNA"

/db\_xref="taxon:32630"

ORIGIN

Query Match 29.9%; Score 26; DB 6; Length 26;  
Best Local Similarity 100.0%; Pred. No. 0.00052;  
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 CATTAATGAGAGATTTTCTGCG 36

Db 1 CATTAATGAGAGATTTTCTGCG 26

## RESULT 10

## AL390237

## LOCUS

## DEFINITION

Human DNA sequence from clone RP11-278J20 on chromosome 6. Contains ESTs, STSs and GSSs. Contains an RBP4 (retinoblastoma-binding protein 4) pseudogene and a KIAA0797 pseudogene, complete sequence.

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

## DIRECT SUBMISSION

## SUBMITTED (11-DEC-2000)

## SANGER CENTRE, HINXON, CAMBRIDGESHIRE,

## CB10 1SA, UK. E-MAIL ENQUIRIES: hunkney@sanger.ac.uk

## REQUESTS: clonerequest@sanger.ac.uk

## ON SEP 13, 2000 THIS SEQUENCE VERSION REPLACED GI:9994265.

## DURING SEQUENCE ASSEMBLY DATA IS COMPARED FROM OVERLAPPING CLONES.

## WHERE DIFFERENCES ARE FOUND THESE ARE ANNOTATED AS VARIATIONS

## TOGETHER WITH A NOTE OF THE OVERLAPPING CLONE NAME. NOTE THAT THE

## VARIATION ANNOTATION MAY NOT BE FOUND IN THE SEQUENCE SUBMISSION

## CORRESPONDING TO THE OVERLAPPING CLONE, AS WE SUBMIT SEQUENCES WITH

## ONLY A SMALL OVERLAP AS DESCRIBED ABOVE.

## THE FOLLOWING ABBREVIATIONS ARE USED TO ASSOCIATE PRIMARY ACCESSION

## NUMBERS GIVEN IN THE FEATURE TABLE WITH THEIR SOURCE DATABASES:

## EMBL, SWISSPROT, TrEMBL, WPI, WORMBASE, INFORMATION

## ON THE WORMBASE DATABASE CAN BE FOUND AT

## HTTP://WWW.SANGER.AC.UK/PROJECTS/C\_ELEGANS/WORMBASE

## THIS SEQUENCE WAS GENERATED FROM PART OF BACTERIAL CLONE CONFLGS OF HUMAN

## CHROMOSOME 6, CONSTRUCTED BY THE SANGER CENTRE CHROMOSOME 6 MAPPING

## GROUP. FURTHER INFORMATION CAN BE FOUND AT

## HTTP://WWW.SANGER.AC.UK/HGP/CH6

## IMPORTANT: THIS SEQUENCE IS NOT THE ENTIRE INSERT OF CLONE

## RP11-278J20 IT MAY BE SHORTER BECAUSE WE SEQUENCE OVERLAPPING

## SECTIONS ONLY ONCE, EXCEPT FOR A 100 BASE OVERLAP.

## THE TRUE LEFT END OF CLONE RP11-278J20 IS AT 1 IN THIS SEQUENCE.

## THE TRUE LEFT END OF CLONE Xbac-55C20 IS AT 140021 IN THIS

## SEQUENCE. THIS SEQUENCE HAS BEEN FINISHED ACCORDING TO SEQUENCE MAP

## CRITERIA AS FOLLOWS. AN ATTEMPT IS MADE TO RESOLVE ALL SEQUENCING

## PROBLEMS, SUCH AS COMPRESSIONS AND REPEATS, BUT NOT NECESSARILY

## WITHIN KNOWN ANNOTATED REPEAT SEQUENCE ELEMENTS. WHERE THE

## SEQUENCE IS AMBIGUOUS, THERE IS AN ANNOTATION USING THE 'UNSURE'

## FEATURE KEY. RP11-278J20 IS FROM THE LIBRARY RP11-11 CONSTRUCTED

## BY THE GROUP OF PIETER DE JONG. FOR FURTHER DETAILS SEE

## HTTP://WWW.CHORI.ORG/BACPAC/HOME.HTM

## VECTOR: pBAC3.6.

## FEATURES

## source

## 1. .140120 Location/Qualifiers

## /organism="Homo sapiens"

## /mol\_type="genomic DNA"

## /db\_xref="taxon:9606"

## /chromosome="6"

## /clone="RP11-278J20"

## /clone\_1fb="RP11-11.1"

## 2. .984

## /note="L1MA3 repeat: matches 4280. .5257 of consensus"

## 985. .1151

## repeat\_region

## 1152. .1401

## repeat\_region

## /note="AluSc repeat: matches 5413. .5568 of consensus"

## 1402. .2072

## repeat\_region



repeat_region	/note="U1MA3 repeat: matches 5568. .6304 of consensus"
repeat_region	/note="U1MA3 repeat: matches 5257. .5277 of consensus"
repeat_region	2279. .2344
repeat_region	/note="r33 copies 2 mer aa 69% conserved"
repeat_region	2560. .2621
misc_feature	/note="r31 copies 2 mer at 72% conserved"
repeat_region	3248. .3463
repeat_region	/note="match: GSS: Em:AQ059990"
repeat_region	3442. .3571
misc_feature	/note="MIR repeat: matches 89. .249 of consensus"
repeat_region	3490. .3608
repeat_region	/note="match: GSS: Em:AQ059990"
repeat_region	5092. .5763
repeat_region	/note="U1PA9 repeat: matches 5178. .5854 of consensus"
repeat_region	5779. .6067
repeat_region	/note="U1PA9 repeat: matches 5859. .6163 of consensus"
repeat_region	7865. .7949
repeat_region	/note="M17H1 repeat: matches 72. .156 of consensus"
repeat_region	8143. .8285
repeat_region	/note="M17H1 repeat: matches 404. .547 of consensus"
repeat_region	8387. .8806
misc_feature	/note="MSB repeat: matches 1. .426 of consensus"
repeat_region	9159. .9604
repeat_region	/note="match: GSS: Em:AQ695871"
repeat_region	9562. .10457
repeat_region	/note="U1MB4 repeat: matches 5290. .6182 of consensus"
repeat_region	11290. .11394
repeat_region	/note="MIR repeat: matches 65. .179 of consensus"
repeat_region	11848. .12266
misc_feature	/note="MSR repeat: matches 3. .426 of consensus"
misc_feature	Complement(12949. .13532)
misc_feature	/note="match: GSS: Em:AQ389585"
misc_feature	Complement(13299. .13630)
misc_feature	/note="match: GSS: Em:B38390"
repeat_region	13542. .14013
repeat_region	/note="match: GSS: Em:AQ719307"
repeat_region	14021. .14444
repeat_region	/note="MSB repeat: matches 1. .426 of consensus"
repeat_region	15546. .15724
repeat_region	/note="MSB20 repeat: matches 4. .177 of consensus"
repeat_region	15842. .15936
repeat_region	/note="U2 repeat: matches 2629. .2748 of consensus"
repeat_region	16125. .16230
repeat_region	/note="U1PA15 repeat: matches 4040. .6151 of consensus"
misc_feature	18459. .18906
misc_feature	/note="match: GSS: Em:AQ240510"
misc_feature	18455. .18819
misc_feature	/note="match: GSS: Em:B31098"
misc_feature	18481. .18817
misc_feature	/note="match: GSS: Em:B31099"
misc_feature	Complement(18659. .18945)
repeat_region	/note="match: GSS: Em:AQ527813"
repeat_region	18856. .19135
repeat_region	/note="A1u0 repeat: matches 5. .282 of consensus"
repeat_region	19273. .19348
repeat_region	/note="MSB5B repeat: matches 2. .78 of consensus"
repeat_region	19357. .19418
repeat_region	/note="MSB92C repeat: matches 494. .546 of consensus"
repeat_region	19402. .19767
repeat_region	/note="MSB67D repeat: matches 71. .436 of consensus"
repeat_region	19768. .21133
repeat_region	/note="U1PA8 repeat: matches 2430. .3809 of consensus"
repeat_region	21135. .23453
repeat_region	/note="U1PA8 repeat: matches 3824. .6163 of consensus"
repeat_region	23488. .23568
repeat_region	/note="MSB67D repeat: matches 5. .88 of consensus"
repeat_region	23852. .23909
repeat_region	/note="MSB53 repeat: matches 1. .58 of consensus"
repeat_region	26262. .26580
repeat_region	/note="A1u0 repeat: matches 1. .311 of consensus"
repeat_region	28100. .28165
repeat_region	/note="MSB9b repeat: matches 457. .553 of consensus"

```

repeat_region      28333. 28656
/notes="AluXc repeat: matches 1. 312 of consensus"
repeat_region      28875. 28908
/notes="MER34 repeat: matches 148. 177 of consensus"
repeat_region      28909. 29004
/notes="LTR29 repeat: matches 4. 100 of consensus"
repeat_region      29005. 29015
/notes="MER34 repeat: matches 21. 148 of consensus"
repeat_region      29817. 30046
/notes="TIGER1 repeat: matches 2175. 2413 of consensus"
repeat_region      30047. 30342
/notes="AluSc repeat: matches 1. 294 of consensus"
repeat_region      30333. 30860
/notes="TIGER1 repeat: matches 1695. 2175 of consensus"
repeat_region      30855. 30900
/notes="TIGER1 repeat: matches 198. 250 of consensus"
repeat_region      30985. 31132
/notes="TIGER1 repeat: matches 23. 166 of consensus"
gene
32392. 33680
/gene="bA278J20.1"
/pseudo
32392. 33680
/gene="bA278J20.1"
/notes="bA278J20.1" (RBP4 (retinoblastoma-binding protein
4) pseudogene)
match: cDNAs: Em:Af073787 Em:Af090306 Em:X71810 Em:U35141
Em:U35143 Em:X74262 Em:X72841 Em:Af057750
match: ESTs: Em:AA534337 Em:AA536652 Em:A1914754
Em:AA040411 Em:AA403752 Em:A1149668 Em:AA312570
Em:AW402945 Em:AA354104 Em:AA313513 Em:AW362873
Em:AA705035 Em:AA263144 Em:AA337765 Em:AA306805
Em:AA070448 Em:AA429422 Em:AW362878 Em:567309 Em:AA853720
Em:AA429422 Em:A1099022 Em:AA359717 Em:AA359673
Em:AL045286 Em:H51237 Em:AA227133 Em:AA310336 Em:AA040303
Em:AA330258 Em:AA331743 Em:H06565 Em:AW86086 Em:AA074922
Em:AA330258 Em:AA080041 Em:AA082962 Em:AA292984
Em:AW386068 Em:H39087 Em:AA281307 Em:AA582542 Em:AA313395
Em:AA730527 Em:AA355615 Em:AA171394 Em:AW367949
Em:AA815284 Em:A1138310 Em:A1060394 Em:N92003 Em:AW403348
Em:D56466 Em:AL038896
match: proteins: Sw:P90917 Sw:Q22468 Sw:Q22469 Sw:P90916
Tr:O94244 Sw:P39984 Sw:O14021 Sw:P43254 Tr:O23774
Sw:O09028 Sw:Q24572 Sw:Q60972 Sw:Q60973 Tr:O98078
Sw:Q22466 Sw:Q22467 Tr:O00809 Sw:P93471 Sw:O16576
Tr:O9Y825 Sw:Q22607"
/pseudo
/codon_start=1
/evidence=not_experimental
35461. 35982
/notes="MER51B repeat: matches 3. 617 of consensus"
repeat_region      36651. 37134
/notes="L2 repeat: matches 2025. 2595 of consensus"
repeat_region      37135. 37451
/notes="AluXc repeat: matches 1. 312 of consensus"
repeat_region      37452. 37603
/notes="L2 repeat: matches 2594. 2745 of consensus"
repeat_region      37683. 37811
/notes="MIR repeat: matches 68. 200 of consensus"
repeat_region      37812. 37845
/notes="L7 copies 2 mer ca 88% conserved"
repeat_region      37889. 38025
/notes="LTR45 repeat: matches 349. 486 of consensus"
repeat_region      40149. 40908
/notes="LMB7 repeat: matches 5406. 6173 of consensus"
repeat_region      40950. 41131
Query Match      24.1%; Score 21; DB 9; Length 140120;
Best Local Similarity 100.0%; Pired. No. 0.18;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```



RESULT 11  
AC026139  
LOCUS  
DEFINITION  
AC026139 142126 bp DNA linear HTG 06-MAY-2001  
SEQUENCE, 23 unordered pieces.  
AC026139  
AC026139.2 GI:9994157  
VERSION  
HTG; HTGS PHASE1; HTGS\_DRAFT.  
KEYWORDS  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 142126)  
Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
Homo sapiens chromosome 6, clone RP11-278J20  
Unpublished  
2 (bases 1 to 142126)  
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,  
Boguslavsky, L., Bouhagalter, B., Brown, A., Burkett, G., Collins, S.,  
Campopiano, A., Castle, A., Choapel, Y., Colangelo, M., Collins, S.,  
Collamore, A., Cooke, P., Dearrellano, K., Dewar, K., Diaz, J. S.,  
Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, M., Gage, D.,  
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,  
Grand-pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L.,  
Howard, J. C., Iliev, I., Johnson, R., Jones, C., Kam, L., Karatas, A.,  
Klein, J., Lacroque, K., Lamazares, R., Landers, T., Lehotzky, J.,  
Levine, R., Lieu, C., Liu, G., Locke, K., MacDonald, P., Marquis, N.,  
McCarthy, M., McEwan, P., McGuirk, A., McKernan, K., McPheters, R.,  
McLirim, J., Meneses, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,  
Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,  
O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N., P.,  
Pisani, C., Pollard, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talama, J.,  
Tessile, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J.,  
Vassiliou, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,  
Young, G., Zainoun, J., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (19-MAR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Sep 8, 2000 this sequence version replaced gi:7264209.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: MIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L8025  
Center clone name: 278\_J 20  
----- Summary Statistics  
Sequencing vector: M13; M77815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 132425 bases at least Q40  
Consensus quality: 137265 bases at least Q30  
Consensus quality: 139029 bases at least Q20  
Insert size: 14000; agarose-fp  
Insert size: 139926; sum-of-contigs  
Quality coverage: 4.5 in Q20 bases; agarose-fp  
Quality coverage: 4.5 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 23 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will  
\* be preserved.  
1  
16449: contig of 16449 bp in length  
16450: gap of 100 bp  
16550: contig of 1288 bp in length  
17837: gap of 100 bp  
17938: contig of 1206 bp in length  
19143: gap of 100 bp  
19243: contig of 1544 bp in length  
20787: gap of 100 bp  
20788: contig of 2434 bp in length  
20888: gap of 100 bp  
22322: contig of 2129 bp in length  
23422: gap of 100 bp  
25551: contig of 2577 bp in length  
25651: gap of 100 bp  
28328: contig of 2232 bp in length  
28328: gap of 100 bp  
30560: contig of 2514 bp in length  
30660: gap of 100 bp  
33173: contig of 3249 bp in length  
33274: gap of 100 bp  
33274: contig of 3249 bp in length  
36523: gap of 100 bp  
36523: contig of 3572 bp in length  
40194: gap of 100 bp  
40195: contig of 5872 bp in length  
40284: gap of 100 bp  
40295: contig of 5298 bp in length  
45593: gap of 100 bp  
45593: contig of 7581 bp in length  
53273: gap of 100 bp  
53273: contig of 7581 bp in length  
53374: gap of 100 bp  
53374: contig of 5872 bp in length  
59245: gap of 100 bp  
59246: gap of 100 bp  
59346: contig of 7157 bp in length  
65032: gap of 100 bp  
65032: contig of 7157 bp in length  
66602: gap of 100 bp  
66602: contig of 8344 bp in length  
74946: gap of 100 bp  
74946: contig of 8344 bp in length  
75047: gap of 100 bp  
75047: contig of 7229 bp in length  
82285: gap of 100 bp  
82285: contig of 7229 bp in length  
82385: gap of 100 bp  
82385: contig of 7932 bp in length  
90378: gap of 100 bp  
90378: contig of 8990 bp in length  
90477: gap of 100 bp  
90477: contig of 8990 bp in length  
95657: gap of 100 bp  
95657: contig of 11247 bp in length  
110815: gap of 100 bp  
110815: contig of 12602 bp in length  
110915: gap of 100 bp  
110915: contig of 12602 bp in length  
123517: gap of 100 bp  
123517: contig of 13104 bp in length  
123617: gap of 100 bp  
123617: contig of 13104 bp in length  
136721: gap of 100 bp  
136721: contig of 5306 bp in length.  
142126: contig of 5306 bp in length.  
Location/Qualifiers  
1. 142126  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/map="6"  
/clone="RP11-278J20"  
/clone\_1b="RP11-278J20 Human Male BAC"  
1. 16449  
/note="assembly\_fragment"  
Clone\_end:SP6  
vector\_side:left  
16550..17837  
/note="assembly\_fragment"  
17938..19143  
/note="assembly\_fragment"  
19243..20787  
/note="assembly\_fragment"  
20788..22322  
/note="assembly\_fragment"  
22322..23422  
/note="assembly\_fragment"  
23422..25551  
/note="assembly\_fragment"  
25551..28328  
/note="assembly\_fragment"  
28328..30559  
/note="assembly\_fragment"



```

misc_feature /note="assembly_fragment"
30660..33173
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
33274..36522
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
36623..40194
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
40295..45592
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
45693..53273
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
53374..59245
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
59346..66502
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
66603..74946
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
75047..82285
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
82386..90377
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
90478..99467
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
99568..110814
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
110915..112516
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
123617..136720
/note="assembly_fragment"
misc_feature /note="assembly_fragment"
136821..142126
/note="assembly_fragment"
clone_end:17
vector_side:right"

```

## ORIGIN

```

Query Match 24.1%; Score 21; DB 2; Length 142126;
Best Local Similarity 100.0%; Pred. No. 0.18;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 51 GGAGGCAAGAGACACTCTGG 71
Db 135317 GGAGGCAAGAGACACTCTGG 135337

```

```

RESULT 12
AC025018/c 159942 bp DNA linear PRI 01-MAY-2002
LOCUS Homo sapiens chromosome 18, clone RP11-320C19, complete sequence.
AC025018
VERSION AC025018.6 GI:20377056
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 159942)
REFERENCE Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Unpublished
JOURNAL 2 (bases 1 to 159942)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Bouckgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collimore,A., Cooke,P., Deatrellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domingo,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heatford,A., Horton,L.,
Howard,J.C., Iliev,I., Johnson,R., Jones,C., Kam,L., Karakas,A.,
Klein,J., Liu,C., Liu,G., Locke,K., Lamazares,R., Landers,T., Lehoczkyl,J.,
Levine,R., Liew,C., Liu,G., Locke,K., MacDonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McKernan,K., McKernan,K., McPheters,R.,
Melgrim,J., Menes,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,

```

```

TITLE JOURNAL
REFERENCE AUTHORS

```

```

TITLE JOURNAL
REFERENCE AUTHORS

```

```

Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Strange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (03-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 159942)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Bouckgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collimore,A.,
Cooke,A., Cooke,P., Deatrellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karakas,A., Kells,C., Labocque,K., Lamazares,R.,
Landers,T., Lehoczkyl,J., Levine,R., Lindblad-Toh,K., Liu,G.,
Maclean,C., MacDonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Melgrim,J., Menes,L.,
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Riese,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnupack,R.,
Seaman,S., Severy,P., Spencer,B., Strange-Thomann,N., Stojanovic,N.,
Straus,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (28-APR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 159942)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Bouckgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collimore,A.,
Cooke,A., Cooke,P., Deatrellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karakas,A., Kells,C., Labocque,K., Lamazares,R.,
Landers,T., Lehoczkyl,J., Levine,R., Lindblad-Toh,K., Liu,G.,
Maclean,C., MacDonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Melgrim,J., Menes,L.,
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Riese,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnupack,R.,
Seaman,S., Severy,P., Spencer,B., Strange-Thomann,N., Stojanovic,N.,
Straus,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (01-MAY-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 1, 2002 this sequence version replaced gi:20336169.
All repeats were identified using RepeatMasker:
Smith, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/MN/RepeatMasker.html
Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
Project Information

```



Center project name: L7797  
Center clone name: 320\_C\_19

FEATURES  
-----  
Location/Qualifiers  
1..159942  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="18"  
/map="18"  
/clone="RP11-320C19"  
/clone.lib="RPCT-11 Human Male BAC"

repeat\_region  
63..203  
/rpt\_family="(GGA)n"  
complement(733..898)  
/rpt\_family="Char11e8"  
2441..3030  
/rpt\_family="L1M3c"  
3037..3241  
/rpt\_family="AluOo"  
3243..3813  
/rpt\_family="L1"  
3814..4126  
/rpt\_family="AluSx"  
4127..4342  
/rpt\_family="L1"  
4343..4545  
/rpt\_family="MER1B"  
complement(4546..4852)  
/rpt\_family="AluSg"  
4853..5013  
/rpt\_family="MER1B"  
5014..6173  
/rpt\_family="L1"  
6186..8358  
/rpt\_family="L1P3"  
complement(8357..8409)  
/rpt\_family="L1P8"  
complement(8410..8556)  
/rpt\_family="L1P"  
complement(8535..8801)  
/rpt\_family="L1P"  
8803..9894  
/rpt\_family="L1P8"  
9896..10358  
/rpt\_family="L1P8"  
10364..12455  
/rpt\_family="L1"  
12456..12747  
/rpt\_family="AluOo"  
12762..13630  
/rpt\_family="(TR)n"  
complement(13146..13422)  
/note="single clone coverage"

unsure  
13615..14094  
/rpt\_family="(TNTAA)n"  
14053..14219  
/rpt\_family="(TR)n"  
14241..14392  
/rpt\_family="L1M1"  
14483..14990  
/rpt\_family="L1M9"  
complement(15265..15543)  
/rpt\_family="AluY"  
16050..16079  
/rpt\_family="(GA)n"  
complement(16838..17311)  
/rpt\_family="AluSc"  
18002..18044  
/rpt\_family="AT\_rich"  
18259..18476  
/rpt\_family="MIR"  
18512..20673

repeat\_region  
complement(20738..20955)  
/rpt\_family="MER56A"  
21395..21417  
/rpt\_family="AT\_rich"  
21958..22317  
/rpt\_family="THE1B"  
22318..23806  
/rpt\_family="THE1B-int"  
23807..24173  
/rpt\_family="THE1B"  
24189..24239  
/rpt\_family="(TCTG)n"  
24244..24281  
/rpt\_family="(CA)n"  
24437..24468  
/rpt\_family="AT\_rich"  
25078..25307  
/rpt\_family="MIR"  
25413..25569  
/rpt\_family="FLAM\_A"  
26749..26790  
/rpt\_family="AT\_rich"  
complement(26905..27421)  
/rpt\_family="L1M1"  
27427..27938  
/rpt\_family="MER66B"

Query Match 24.1% Score 21; DB 9; Length 159942;  
Best Local Similarity 100.0%; Pred. No. 0.18;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 GGAGGCAAGAGCACTCTGG 71  
Db 91013 GGAGGCAAGAGCACTCTGG 90993

RESULT 13  
AL512427  
LOCUS  
DEFINITION Human DNA sequence from clone RP11-325M4 on chromosome 6, complete  
sequence.  
ACCESSION AL512427  
VERSION AL512427.10 GI:14270153  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homindaee; Homo.  
1 (bases 1 to 201657)  
REFERENCE  
AUTHORS Williams,S.  
TITLE Submitted (25-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
COMMENT  
On May 31, 2001 this sequence version replaced GI:13235016.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em: EMBL; Sw:  
SWISSPROT; Tr: TREMBL; Wp: WORMBASE; information on the WORMBASE  
database can be found at



[http://www.sanger.ac.uk/Projects/C\\_elegans/wormed](http://www.sanger.ac.uk/Projects/C_elegans/wormed) This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6> RP11-325M4 is from the library RPCT-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>  
 VECTOR: pBACe3.6

## FEATURES

```

source
1..201657
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP11-325M4"
/clone_id="RPC1-11.2"
7..696
/misc_feature
/note="match: GSS: Em:AQ536961"
14..725
/misc_feature
/note="match: GSS: Em:AQ395017"
19..429
/misc_feature
/note="match: GSS: Em:AQ390879"
157..2551
/repeat_region
/note="TIGER1 repeat: matches 11..2418 of consensus"
2582..2665
/repeat_region
/note="42 copies 2 mer ta 63% conserved"
3974..4332
/repeat_region
/note="WTA1 repeat: matches 1..363 of consensus"
4333..4378
/note="23 copies 2 mer gt 82% conserved"
complement(4857..5207)
/misc_feature
/note="match: GSS: Em:AQ455938"
complement(4960..5518)
/misc_feature
/note="match: GSS: Em:AQ375389"
complement(4963..5512)
/misc_feature
/note="match: GSS: Em:AQ377395"
complement(5221..5487)
/misc_feature
/note="match: GSS: Em:AQ537274"
5519..6021
/misc_feature
/note="match: GSS: Em:AQ525186"
8849..8904
/repeat_region
/note="MIR repeat: matches 196..252 of consensus"
8926..9202
/repeat_region
/note="11P5 repeat: matches 5866..6142 of consensus"
9203..9537
/repeat_region
/note="11P5 repeat: matches 5536..5873 of consensus"
9553..9709
/repeat_region
/note="MIR repeat: matches 2..155 of consensus"
10469..10786
/repeat_region
/note="AluB repeat: matches 1..306 of consensus"
12502..12885
/repeat_region
/note="WTA1B repeat: matches 1..390 of consensus"
13350..13494
/repeat_region
/note="MIR repeat: matches 6..154 of consensus"
complement(13695..14217)
/misc_feature
/note="match: GSS: Em:AQ732279"
complement(13702..14228)
/misc_feature
/note="match: GSS: Em:AQ892616"
complement(13736..14178)
/misc_feature
/note="match: GSS: Em:AQ571323"
complement(13743..14205)
/misc_feature
/note="match: GSS: Em:AQ808008"
14570..14850
/misc_feature
/note="match: GSS: Em:AQ192455"
15424..15596
/repeat_region
/note="11M1 repeat: matches -1388..-1207 of consensus"
15639..15710
/repeat_region
/note="AluU/FRM repeat: matches 223..294 of consensus"
15713..16235
/repeat_region
/note="11R9 repeat: matches 14..625 of consensus"
16328..16468
/repeat_region

```







Db 15534 AGTGAGATTCTTGACCTG 15515

RESULT 16  
ACI03355/c  
LOCUS ACI03355  
DEFINITION Mus musculus chromosome 14 clone RP24-482C10 map 14, \*\*\* SEQUENCING  
IN PROGRESS \*\*\* 3 unordered pieces.

ACCESSION  
ACI03355  
VERSION HTG: HTGS PHASE1; HTGS\_FULFILL; HTGS\_ACTIVERIN.  
KEYWORDS Mus musculus (house mouse)  
SOURCE Mus musculus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
Birken, B., Nusbaum, C. and Lander, E.  
1 (bases 1 to 195708)  
Mus musculus chromosome 14, clone RP24-482C10  
Unpublished  
2 (bases 1 to 195708)

AUTHORS  
Birken, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B., Brown, A., Camarata, U., Campoliano, A., Chang, J., Chazaro, B., Choepel, Y., Collangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J., Dodge, S., Fero, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Coyette, M., Graham, L., Grand-pierre, N., Hages, B., Heaford, A., Horton, L., Hulme, W., Illiev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Lacroque, K., Lamaszates, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McGowan, P., McKernan, K., McPheeters, R., Meldrim, J., Menais, L., Minova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nord, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schopack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strause, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travers, N., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE  
Submitted (24-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 195708)

REFERENCE  
Birken, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, M., Arachchi, H.M., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J., S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Fero, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-pierre, N., Hages, B., Heaford, A., Horton, L., Hulme, W., Illiev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, X., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrim, J., Menais, L., Minova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Nord, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schupack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE  
Direct Submission

JOURNAL  
Submitted (11-DEC-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Dec 11, 2003 this sequence version replaced gi:138454415.

COMMENT  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L18635  
Center clone name: 482\_C\_10  
-----  
\* NOTES: This is a 'working draft' sequence. It currently  
\* consists of 3 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 146464: contig of 146464 bp in length  
\* 146465 146564: gap of 100 bp  
\* 146565 192519: contig of 45955 bp in length  
\* 192520 192619: gap of 100 bp  
\* 192620 195708: contig of 3089 bp in length.  
Location/Qualifiers  
1. 195708  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="14"  
/map="14"  
/clone="RP24-482C10"  
/clone\_1b="PFC1-24 Male Mouse BAC"

ORIGIN  
Query Match 23.0%; Score 20; DB 2; Length 195708;  
Best Local Similarity 100.0%; Fred. No. 0.7;  
Matches 20; Conservative 0; Mismatches 0; Gaps 0;  
32 TCTGGCCTGAGACTTGACG 51  
Db 134902 TCTGGCCTGAGACTTGACG 134883

RESULT 17  
AC073278/c  
LOCUS AC073278 200807 bp DNA linear PRI 08-OCT-2003  
DEFINITION Homo sapiens BAC clone RP11-66118 from 7, complete sequence.  
ACCESSION AC073278  
VERSION AC073278.9 GI:15145619  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homindaes; Homo.  
1 (bases 1 to 200807)  
Toward a complete human genome sequence  
Sullivan, J.E. and Wilson, R.  
2 (bases 1 to 200807)  
Toward a complete human genome sequence  
Genome Res. 8 (11), 1097-1108 (1998)  
9847074  
PUBMED 99063792  
TITLE  
JOURNAL  
Submitted (12-JUN-2000) Genome Sequencing Center, Washington University School of Medicine, 444 Forest Park Parkway, St. Louis, MO 63108, USA  
4 (bases 1 to 200807)  
Waterson, R.H.  
Direct Submission  
JOURNAL  
Submitted (09-AUG-2001) Genome Sequencing Center, Washington



University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
REFERENCE  
5 (bases 1 to 200807)  
AUTHORS  
Waterson,R.  
JOURNAL  
Submitted (09-JUN-2002) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
6 (bases 1 to 200807)  
AUTHORS  
Wilson,R.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (08-OCT-2003) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Aug 9, 2001 this sequence version replaced gi:14476026.  
COMMENT  
-----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu>  
Contact: [sapienewatson.wustl.edu](mailto:sapienewatson.wustl.edu)  
-----  
Summary Statistics  
Center project name: H\_NH0661L08  
-----

NOTICE: This sequence may not represent the entire insert of this  
clone. It may be shorter because we only sequence overlapping  
clone sections once, or longer because we provide a small overlap  
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate  
chemistry, or covered by high quality data (i.e., phred quality >=  
30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by sequence  
from more than one subclone; and the assembly was confirmed by  
restriction digest.

MAPPING INFORMATION:  
The sequence of this clone was established as part of a mapping and  
sequencing collaboration between the NHGRI Chromosome 7 Mapping  
Project (Eric D. Green, Director), John D. McPherson in the  
Department of Genetics (Washington University), and the Washington  
University Genome Sequencing Center. For additional information  
about the map position of this sequence, see  
<http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send  
<mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu>

SOURCE INFORMATION:  
The RPCI-11 human BAC library was made from the blood of one male  
donor, as described by Osoegawa,K., Moon,P.Y., Zhao,B., Frenken,B.,  
Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved  
approach for construction of bacterial artificial chromosome  
libraries. Genomics 51:1-8. The clone may be obtained either from  
Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong  
and coworkers at <http://www.chori.org>  
VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:  
The clone sequenced to the left is RP11-627B12 the clone sequenced  
to the right is RP5-1102A12, 2000 bp overlap. Actual start of this  
clone is at base position 1 of RP11-661L8.

## FEATURES

SOURCE  
1. 200807  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="7"  
/map="17"  
/clone="RP11-661L8"  
/clone\_lib="RPCI-11"  
38..98  
/rpt\_family="CRI"  
repeat\_region  
645..935  
/rpt\_family="Alu"  
repeat\_region  
1113..1410

repeat\_region /rpt\_family="Alu"  
1858..2040  
repeat\_region /rpt\_family="Alu"  
2202..2308  
repeat\_region /rpt\_family="Alu"  
2392..2428  
repeat\_region /rpt\_family="L2"  
2615..3035  
repeat\_region /rpt\_family="Alu"  
3036..3197  
repeat\_region /rpt\_family="MIR"  
3357..3458  
repeat\_region /rpt\_family="C-rich"  
3456..3785  
repeat\_region /rpt\_family="Alu"  
3982..4295  
repeat\_region /rpt\_family="Alu"  
4302..4377  
repeat\_region /rpt\_family="L1"  
4518..4790  
repeat\_region /rpt\_family="Alu"  
4791..4812  
repeat\_region /rpt\_family="(CA)n"  
4852..4934  
repeat\_region /rpt\_family="L1"  
4927..5144  
repeat\_region /rpt\_family="L1"  
5145..5427  
repeat\_region /rpt\_family="Alu"  
5428..5796  
repeat\_region /rpt\_family="L1"  
5882..5915  
repeat\_region /rpt\_family="Alu"  
6061..6383  
repeat\_region /rpt\_family="L1"  
6384..6430  
repeat\_region /rpt\_family="(TG)n"  
6431..6684  
repeat\_region /rpt\_family="Alu"  
6685..6723  
repeat\_region /rpt\_family="L1"  
6795..6825  
repeat\_region /rpt\_family="(CA)n"  
6826..6920  
repeat\_region /rpt\_family="(TA)n"  
6932..7226  
repeat\_region /rpt\_family="Alu"  
7245..7370  
repeat\_region /rpt\_family="Alu"  
7574..7599  
repeat\_region /rpt\_family="AT-rich"  
7868..8008  
repeat\_region /rpt\_family="Alu"  
8031..8053  
repeat\_region /rpt\_family="BC200"  
8054..8418  
repeat\_region /rpt\_family="MIR"  
8419..8543  
repeat\_region /rpt\_family="BC200"  
8671..8796  
repeat\_region /rpt\_family="MIR"  
9175..9453  
repeat\_region /rpt\_family="Alu"  
9481..9767  
repeat\_region /rpt\_family="Alu"  
9772..9899  
repeat\_region /rpt\_family="Alu"  
9900..9936  
repeat\_region /rpt\_family="(CA)n"  
9937..10100  
repeat\_region /rpt\_family="Alu"  
10413..10450  
repeat\_region /rpt\_family="(TTCTC)n"



```

repeat_region      10451..10753
                    /rpt_family="Alu"
repeat_region      11141..11472
                    /rpt_family="L2"
repeat_region      11473..11745
                    /rpt_family="Alu"
repeat_region      11746..11965
                    /rpt_family="L2"
repeat_region      12550..12690
                    /rpt_family="Alu"
repeat_region      13003..13140
                    /rpt_family="L1"
repeat_region      13141..13441
                    /rpt_family="Alu"
repeat_region      13442..13692
                    /rpt_family="L1"
repeat_region      13693..13979
                    /rpt_family="Alu"
repeat_region      13980..14002
                    /rpt_family="L1"
repeat_region      14003..14312
                    /rpt_family="Alu"
repeat_region      14313..14416
                    /rpt_family="L1"
repeat_region      14559..14921

Query Match      23.0% Score 20; DB 9; Length 200807;
Best Local Similarity 100.0%; Pred. NO. 0.7;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      23 TGAGATTTTTCGGCCTGAG 42
DB      59425 TGAGATTTTTCGGCCTGAG 59406

RESULT 18
AC141733
LOCUS      AC141733
DEFINITION Apis mellifera clone CH224-58F20, WORKING DRAFT SEQUENCE, 71
ACCESSION AC141733
VERSION    AC141733.1 GI:29123917
KEYWORDS   HTG; HTGS PHASE1; HTGS DRAFT.
SOURCE     Apis mellifera (honeybee)
ORGANISM   Apis mellifera
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea;
            Apidae; Apis.
            1 (bases 1 to 223920)
Muzny,D.M., Adams,C., Adio-Oduola,B., All-oman,F., Allen,C.,
Albrooks,S.L., Amaral-Tunget,H.C., Are,J.R., Ayala,M., Banks,T.,
Barbieri,J., Benton,U., Bimaga,K., Blankenburg,K., Bonin,D.,
Bouck,J., Bowls,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Doutwaite,K.T., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Eamhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falle,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Fratcz,P.,
Gabriel,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.B.,
Jaccoson,B., Jia,Y., Johnson,R., Jolivet,S., Joudan,S.,
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtharge,O., Lieu,C., Liu,J., Liu,W., Louisedge,H.,
Lozago,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Maesey,E., Mawney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokwenwo,S., Ogih,M., Okumura,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,X.,
Rives,M., Rojas,A., Rojudoan,I., Rolle,M., Ruiz,S., Saverly,G.,
Scherer,S., Scott,G., Shen,H., Shooshbari,N., Slason,I.,
Sodergren,E., Sonalke,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Swatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,R., Thomas,S.,
Uman,K., Vaequez,L., Vera,V., Villalobos,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,
Williams,G., Williamson,A., Wleczky,R., Wooten,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

Direct Submission
Unpublished
2 (bases 1 to 223920)
Worley,K.C.
Submitted (19-MAR-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center project name: AMER
Center clone name: CH224-58F20

----- Summary Statistics
Sequencing vector: Plasmid,
Chemistry: Dye-terminator Big Dye 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 208891 bases at least Q40
Consensus quality: 217674 bases at least Q30
Consensus quality: 223660 bases at least Q20
Estimated insert size: 203635; sum-of-ctcontigs estimation
Quality coverage: 3x in Q20 bases; sum-of-ctcontigs estimation

-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a "working draft" sequence. It currently
* consists of 71 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1
1287 1286: contig of 1286 bp in length
1387 1386: gap of unknown length
2587 2587: contig of 1201 bp in length
2588 2687: gap of unknown length
2688 3873: contig of 1186 bp in length
3874 3973: gap of unknown length
3974 5300: contig of 1327 bp in length
5301 5400: gap of unknown length
5401 6899: contig of 1499 bp in length
6900 6999: gap of unknown length
7000 8498: contig of 1499 bp in length
8499 8598: gap of unknown length
8599 9833: contig of 1235 bp in length
9834 9934: gap of unknown length
9934 11085: contig of 1152 bp in length
11086 11185: gap of unknown length
11186 12642: contig of 1457 bp in length
12643 12742: gap of unknown length
12743 13939: contig of 1197 bp in length
13940 14039: gap of unknown length
14040 15762: contig of 1723 bp in length

```



15763 15862: gap of unknown length  
 \* 15863 17042: contig of 1180 bp in length  
 \* 17043 17142: gap of unknown length  
 \* 17143 18188: contig of 1046 bp in length  
 \* 18189 18288: gap of unknown length  
 \* 18289 19305: contig of 1017 bp in length  
 \* 19306 19405: gap of unknown length  
 \* 19406 20891: contig of 1486 bp in length  
 \* 20892 22172: contig of 1181 bp in length  
 \* 22173 22272: gap of unknown length  
 \* 22273 23607: contig of 1335 bp in length  
 \* 23608 23707: gap of unknown length  
 \* 23708 25369: contig of 1662 bp in length  
 \* 25370 25469: gap of unknown length  
 \* 25470 26735: contig of 1266 bp in length  
 \* 26736 26835: gap of unknown length  
 \* 26836 28039: contig of 1264 bp in length  
 \* 28100 28199: gap of unknown length  
 \* 28200 29220: contig of 1021 bp in length  
 \* 29221 29320: gap of unknown length  
 \* 29321 31314: contig of 1994 bp in length  
 \* 31315 31414: gap of unknown length  
 \* 31415 32562: contig of 1148 bp in length  
 \* 32563 32662: gap of unknown length  
 \* 32663 34125: contig of 1463 bp in length  
 \* 34126 34225: gap of unknown length  
 \* 34226 35896: contig of 1671 bp in length  
 \* 35897 35996: gap of unknown length  
 \* 35997 37348: contig of 1352 bp in length  
 \* 37349 37448: gap of unknown length  
 \* 37449 39365: contig of 1917 bp in length  
 \* 39366 39465: gap of unknown length  
 \* 39466 40958: contig of 1493 bp in length  
 \* 40959 41058: gap of unknown length  
 \* 41059 42440: contig of 1382 bp in length  
 \* 42441 42540: gap of unknown length  
 \* 42541 43967: contig of 1427 bp in length  
 \* 43968 44067: gap of unknown length  
 \* 44068 45739: contig of 1672 bp in length  
 \* 45740 45839: gap of unknown length  
 \* 45840 47140: contig of 1301 bp in length  
 \* 47141 47240: gap of unknown length  
 \* 47241 48858: contig of 1718 bp in length  
 \* 48859 49058: gap of unknown length  
 \* 49059 51365: contig of 2307 bp in length  
 \* 51366 51465: gap of unknown length  
 \* 51466 53056: contig of 1591 bp in length  
 \* 53057 53156: gap of unknown length  
 \* 53157 54710: contig of 1554 bp in length  
 \* 54711 54810: gap of unknown length  
 \* 54811 56040: contig of 1230 bp in length  
 \* 56041 56140: gap of unknown length  
 \* 56141 58307: contig of 2167 bp in length  
 \* 58308 58407: gap of unknown length  
 \* 58408 60071: contig of 1664 bp in length  
 \* 60072 60171: gap of unknown length  
 \* 60172 61494: contig of 1323 bp in length  
 \* 61495 61594: gap of unknown length  
 \* 61595 63574: contig of 1980 bp in length  
 \* 63575 63674: gap of unknown length  
 \* 63675 64825: contig of 1151 bp in length  
 \* 64826 64925: gap of unknown length  
 \* 64927 66277: contig of 1352 bp in length  
 \* 66278 66377: gap of unknown length  
 \* 66378 68973: gap of 2596 bp in length  
 \* 68974 69073: gap of unknown length  
 \* 69074 71541: contig of 2466 bp in length  
 \* 71542 73546: contig of 1905 bp in length  
 \* 73547 73647: gap of unknown length  
 \* 73647 75530: contig of 1884 bp in length  
 \* 75531 75630: gap of unknown length

Query Match 23.0%; Score 20; DB 2; Length 223920;  
 Best Local Similarity 100.0%; Pred. No. 0.69;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 15 AATTGAAGTCAGATTTCCT 34  
 Db 106395 AATTGAAGTCAGATTTCCT 106414  
 RESULT 19  
 AC141730/c  
 LOCUS  
 DEFINITION  
 AC141730 274341 bp DNA linear HTG 19-MAR-2003  
 Apis mellifera clone CH224-58F21, WORKING DRAFT SEQUENCE, 90  
 unordered pieces.  
 AC141730  
 VERSION  
 HTG; HTGS PHASE1; HTGS DRAFT.  
 KEYWORDS  
 Apis mellifera (honeybee)  
 ORGANISM  
 Apis mellifera  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea;  
 Apidae; Apis.  
 1 (bases 1 to 274341)  
 REFERENCE  
 AUTHORS  
 Mazyu,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
 Alshrocks,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T.,  
 Barbara,J., Benton,J., Bimge,K., Blankenburg,K., Bonnin,D.,  
 Bouck,J., Bowls,S., Blevins,M., Brown,E., Brown,M., Bryant,N.P.,  
 Bukey,C., Burck,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
 Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinn,H.H.,  
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.U.,  
 Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
 Falls,T., Ferraguno,D., Flagg,N., Ford,J., Foster,P., Franz,P.,  
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
 Gottell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,  
 Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,  
 Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,  
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,  
 Kratochvic,J., Kireeshi,A., Landry,N., Leal,B., Levele,L.C., Lewis,L.,  
 Li,J., Li,Z., Licharge,O., Lien,C., Liu,J., Liu,W., Louisedge,H.,  
 Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
 Maheshwari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E.,  
 Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,  
 Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,  
 Moser,M., Neal,D., Newton,U., Newton,N., Nguyen,A., Nguyen,N.,  
 Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguth,M., Okwuonun,G.,  
 Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
 Peters,L., Pickens,R., Primus,E., Pu,L., Quiles,M., Ren,Y.,  
 Rives,M., Rojas,A., Rojudoan,I., Rolle,M., Ruiz,S., Saverly,C.,



Scherer, S., Scott, G., Shen, H., Shooshari, N., Sisson, I.,  
Sodergren, E., Sotak, T., Sparks, A., Stanley, H., Stone, H.,  
Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
Tansley, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
Uemami, K., Vasquez, L., Vera, V., Villalón, D., Vinsom, R., Wang, Q.,  
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
Williams, G., Williamson, A., Wlezyk, R., Wooden, S., Worley, K.,  
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
Weinstock, G., and Gibbs, R.

Unpublished  
Direct Submission  
2 (bases 1 to 274341)  
Mortley, K.C.

Submitted (19-MAR-2003) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

----- Genome Center -----  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
----- Project Information -----  
Center project name: AMED  
Center clone name: CH224-58P21  
----- Summary Statistics -----  
Sequencing vector: Plasmid  
Chemistry: Dye-terminator Big Dye 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 265284 bases at least Q40  
Consensus quality: 272666 bases at least Q30  
Consensus quality: 278765 bases at least Q20  
Estimated insert size: 242717; sum-of-contigs estimation  
Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

----- NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
NOTE: This is a 'working draft' sequence. It currently  
consists of 90 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

1 1269: contig of 1269 bp in length  
1 1270 1369: gap of unknown length  
\* 1370 2530: contig of 1161 bp in length  
\* 2531 2630: gap of unknown length  
\* 2631 3633: contig of 1003 bp in length  
\* 3634 3733: gap of unknown length  
\* 3734 4734: contig of 1001 bp in length  
\* 4735 4834: gap of unknown length  
\* 4835 6074: contig of 1240 bp in length  
\* 6075 6174: gap of unknown length  
\* 6175 7362: contig of 1188 bp in length  
\* 7363 7462: gap of unknown length  
\* 7463 8561: contig of 1099 bp in length  
\* 8562 8661: gap of unknown length  
\* 8662 10216: contig of 1555 bp in length  
\* 10217 10317: gap of unknown length  
\* 10318 11411: contig of 1095 bp in length  
\* 11412 11511: gap of unknown length  
\* 11512 13018: contig of 1507 bp in length  
\* 13019 13118: gap of unknown length  
\* 13119 14393: contig of 1275 bp in length  
\* 14394 14493: gap of unknown length  
\* 14494 15606: contig of 1113 bp in length  
\* 15607 15706: gap of unknown length  
\* 15707 16994: contig of 1288 bp in length  
\* 16995 17094: gap of unknown length  
\* 17095 18584: contig of 1490 bp in length  
\* 18585 18684: gap of unknown length

18685 19727: contig of 1043 bp in length  
\* 19728 19827: gap of unknown length  
\* 19828 21324: contig of 1497 bp in length  
\* 21325 21424: gap of unknown length  
\* 21425 22440: contig of 1016 bp in length  
\* 22441 22540: gap of unknown length  
\* 22541 23764: contig of 1224 bp in length  
\* 23765 23864: gap of unknown length  
\* 23865 25004: contig of 1140 bp in length  
\* 25005 25104: gap of unknown length  
\* 25105 26501: contig of 1397 bp in length  
\* 26502 26601: gap of unknown length  
\* 26602 27634: contig of 1033 bp in length  
\* 27635 27734: gap of unknown length  
\* 27735 28872: contig of 1138 bp in length  
\* 28873 28972: gap of unknown length  
\* 28973 30355: contig of 1383 bp in length  
\* 30356 30455: gap of unknown length  
\* 30456 31483: contig of 1028 bp in length  
\* 31484 31583: gap of unknown length  
\* 31584 32937: contig of 1354 bp in length  
\* 32938 33037: gap of unknown length  
\* 33038 34788: contig of 1751 bp in length  
\* 34789 34888: gap of unknown length  
\* 34889 36164: contig of 1276 bp in length  
\* 36165 36264: gap of unknown length  
\* 36265 37365: contig of 1101 bp in length  
\* 37366 37465: gap of unknown length  
\* 37466 38815: contig of 1350 bp in length  
\* 38816 38916: gap of unknown length  
\* 38917 40370: contig of 1355 bp in length  
\* 40371 41659: contig of 1289 bp in length  
\* 41660 41759: gap of unknown length  
\* 41760 43462: contig of 1703 bp in length  
\* 43463 43562: gap of unknown length  
\* 43563 44708: contig of 1146 bp in length  
\* 44709 44808: gap of unknown length  
\* 44809 46764: contig of 1956 bp in length  
\* 46765 46864: gap of unknown length  
\* 46865 48291: contig of 1427 bp in length  
\* 48292 48391: gap of unknown length  
\* 48392 49854: contig of 1463 bp in length  
\* 49855 49954: gap of unknown length  
\* 49955 51846: contig of 1892 bp in length  
\* 51847 51946: gap of unknown length  
\* 51947 54235: contig of 2289 bp in length  
\* 54236 55378: gap of unknown length  
\* 55379 55478: contig of 1043 bp in length  
\* 55479 57045: gap of unknown length  
\* 57046 57145: contig of 1567 bp in length  
\* 57146 58809: contig of 1664 bp in length  
\* 58810 58909: gap of unknown length  
\* 58910 60440: contig of 1531 bp in length  
\* 60441 60540: gap of unknown length  
\* 60541 62126: contig of 1586 bp in length  
\* 62127 62226: gap of unknown length  
\* 62227 63825: contig of 1599 bp in length  
\* 63826 63925: gap of unknown length  
\* 63926 65418: contig of 1493 bp in length  
\* 65419 65518: gap of unknown length  
\* 65519 66665: contig of 1147 bp in length  
\* 66666 66765: gap of unknown length  
\* 66766 68207: contig of 1442 bp in length  
\* 68208 69732: gap of unknown length  
\* 69733 69832: contig of 1425 bp in length  
\* 69833 71212: gap of unknown length  
\* 71213 71312: contig of 1380 bp in length  
\* 71313 72563: contig of 1251 bp in length  
\* 72564 72663: gap of unknown length  
\* 72664 74224: contig of 1561 bp in length



```

*      74325      74324: gap of unknown length
*      74325      75580: contig of 1256 bp in length
*      75581      75580: gap of unknown length
*      75681      77306: contig of 1526 bp in length
*      77207      77306: gap of unknown length
*      77307      78998: contig of 1692 bp in length
*      78999      79098: gap of unknown length
*      79099      80808: contig of 1710 bp in length
*      80809      80908: gap of unknown length
*      80909      84064: contig of 3156 bp in length
*      84065      84164: gap of unknown length

```

```

Query Match      23.0%; Score 20; DB 2; Length 274341;
Best Local Similarity 100.0%; Pred. No. 0.67;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      15 AATTGAAGTGAATTTTCT 34
      |||||
Db      65369 AATTGAAGTGAATTTTCT 65350

```

Search completed: March 13, 2004, 08:36:36  
 Job time : 2090 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 13, 2004, 00:39:18 / Search time 302 seconds  
(without alignments)  
1223.819 Million cell updates/sec

Title: US-09-966-880A-11  
Perfect score: 87  
Sequence: 1 agagaacacatcatatga.....ctgacaccatcatgacag 87

Scoring table: OLIGO\_NTC  
Gapop 60.0, Gapext 60.0

Searched: 3373863 seqs, 2124099041 residues

Word size: 20

Total number of hits satisfying chosen parameters: 10

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database: N\_Geneseq\_29Jan04:\*  
1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002s:\*  
7: geneseqn2003as:\*  
8: geneseqn2003bs:\*  
9: geneseqn2003cs:\*  
10: geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	87	100.0	87	3 AAC55315	AAC55315 Human act
2	87	100.0	2818	3 AAC55312	AAC55312 Human act
3	87	100.0	5514	3 AAC55313	AAC55313 Human act
4	87	100.0	11204	3 AAC55339	AAC55339 Human act
5	87	100.0	11204	6 ABS73286	ABS73286 DNA encod
6	84	96.6	1543	7 ABX05468	ABX05468 Human nov
7	84	96.6	2791	6 ABS73287	ABS73287 DNA encod
8	84	96.6	2791	6 ABS73288	ABS73288 DNA encod
9	26	29.9	26	3 AAC55335	AAC55335 Human act
10	20	23.0	20	8 ADB88945	ADB88945 Activatio

ALIGNMENTS

RESULT 1  
AAC55315  
XX AAC55315 standard; DNA; 87 BP.  
XX  
AC AAC55315;  
XX  
DT 05-FEB-2001 (first entry)  
XX

DE Human activation-induced cytidine deaminase exon 1 SEQ ID NO:11.  
KW Activation-induced cytidine deaminase; AID; cytidine deaminase;  
KW immune related disease; allergy; allergic disease; antiallergic;  
KW antianaemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;  
KW gene therapy; B cell associated immune system disorder; food allergy;  
KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
KW IGA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;  
KW drug allergy; allergic rhinitis; Rosen disease; Digeorge disease; AIDS;  
KW ataxia telangiectasia; common variable immunodeficiency disorder;  
KW major histocompatibility class II deficiency disease;  
KW auto immunodeficiency syndrome; Igg subclasses selection disorder; ds.

OS Homo sapiens.  
XX  
XX WO200058480-A1.  
XX  
XX 05-OCT-2000.  
XX  
XX 28-MAR-2000; 2000WO-JP001918.  
XX  
XX 29-MAR-1999; 99JP-00087192.  
XX  
XX 24-JUN-1999; 99JP-00178999.  
XX  
XX 27-DEC-1999; 99JP-00371382.

PA (NTSB ) JAPAN TOBACCO INC.  
PA (HONJ/) HONJO T.

PI Honjo T, Muramatsu M;  
XX  
XX WPI, 2000-611715/58.  
XX

PT Nucleic acid encoding activation induced cytidine deaminase, useful as a  
PT target for drug development for immune-related diseases including  
PS allergies.

PS Claim 18; Page 150; 174pp; Japanese.

CC The present invention describes an activation-induced cytidine deaminase  
CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and has  
CC cytidine activity similar to APOBEC-1. AID has antiallergic, antianaemic,  
CC antiasthmatic, ophthalmological, anti-HIV and dermatological activities,  
CC and can be used in gene therapy. AID polynucleotides are useful in  
CC methods for identifying drugs for the treatment of B cell associated  
CC immune system disorders, immunodeficiency diseases and allergies, such as  
CC immunoglobulin A (IGA) deficiency disease, IGA nephritis, gamma-  
CC globulinaemia, atopic dermatitis, allergic colitis, asthma, food allergy,  
CC drug allergy, allergic rhinitis, Rosen disease, Digeorge disease, ataxia  
CC telangiectasia, common variable immunodeficiency disorder, MHC (major  
CC histocompatibility class) class II deficiency disease, AIDS (auto  
CC immunodeficiency syndrome), elevated Igg disorder, and Igg subclasses  
CC selection disorder. The DNA sequences encoding AID may be used for gene  
CC therapy and the antibodies to the AID protein may be used for diagnosis  
CC and treatment of these disorders. The present sequence represents the  
CC exon 1 genomic DNA sequence of human AID

XX Sequence 87 BP; 28 A; 17 C; 23 G; 19 T; 0 U; 0 Other;  
XX  
XX Query Match 100.0%; Score 87; DB 3; Length 87;  
XX Best local similarity 100.0%; Pred. No. 8.6e-37;  
XX Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAGAACCATCATTAATTAAGTGAATTTTCTGCGCTGAGACTTGACAGGAGCAAGA 60  
DB 1 AGAGAACCATCATTAATTAAGTGAATTTTCTGCGCTGAGACTTGACAGGAGCAAGA 60  
QY 61 AGACACTCTGGACACCACTATGACAG 87  
DB 61 AGACACTCTGGACACCACTATGACAG 87

RESULT 2  
AAC55312



ID	AA655312 standard; cDNA; 2818 BP.
XX	AA655312;
XX	
DT	05-FEB-2001 (first entry)
DE	Human activation-induced cytidine deaminase encoding cDNA seq ID NO:7.
XX	
KW	Activation-induced cytidine deaminase; AID; cytidine deaminase;
KW	immune related disease; allergy; allergic disease; antiallergic;
KW	antianaemic; antiasthmatic; ophthalmological; anti-HIV; dermatologic;
KW	gene therapy; B cell associated immune system disorder; food allergy;
KW	immunodeficiency disease; immunoglobulin A deficiency disease; ascima;
KW	Iga nephritis; gamma-globulinemia; atopic dermatitis; allergic colitis;
KW	drug allergy; allergic rhinitis; Rosen disease; DiGeorge disease; AIDS;
KW	ataxia telangiectasia; common variable immunodeficiency disorder;
KW	major histocompatibility class II deficiency disease;
XX	auto immunodeficiency syndrome; Igg subclass selection disorder; sa.
OS	Homo sapiens.
FH	Key Location/Qualifiers
FT	CDS .676
FT	/tag= a
FT	/product= "activation-induced cytidine deaminase"
XX	
EN	WC200058480-AI.
PD	
XX	05-OCT-2000.
PF	28-MAR-2000; 2000WO-JP001918.
XX	
ER	29-MAR-1999; 99JP-000877192.
PR	24-JUN-1999; 99JP-00178999.
XX	
PR	27-DEC-1999; 99JP-00371382.
PA	(NIBS) JAPAN TOBACCO INC.
PA	(HONJ/) HONJO T.
XX	
FI	Honjo T, Muramatsu M;
XX	
DR	WPI; 2000-611715/58.
DR	P-PSTD3; AAB24198.
PT	Nucleic acid encoding activation induced cytidine deaminase, useful as a
PT	target for drug development for immune-related diseases including
PT	allergies.
PS	
PS	Claim 3; Page 135-139; 174pp; Japanese.
CC	The present sequence encodes human activation-induced cytidine deaminase
CC	(AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and has
CC	cytidine activity similar to APOBEC-1. AID has antiallergic, antianaemic,
CC	antiasthmatic, ophthalmological, anti-HIV and dermatological activities,
CC	and can be used in gene therapy. AID polynucleotides are useful in
CC	methods for identifying drugs for the treatment of B cell associated
CC	immune system disorders, immunodeficiency diseases and allergies, such as
CC	immunoglobulin A (IGA) deficiency disease, IGA nephritis, gamma-
CC	globulinemia, atopic dermatitis, allergic colitis, asthma, food allergy,
CC	drug allergy, allergic rhinitis, Rosen disease, DiGeorge disease, ataxia
CC	telangiectasia, common variable immunodeficiency disorder, MHC (major
CC	histocompatibility class II deficiency disease, AIDS (auto
CC	immunodeficiency syndrome), elevated IGE disorder, and IGG subclass
CC	selection disorder. The DNA sequences encoding AID may be used for gene
CC	therapy and the antibodies to the AID protein may be used for diagnosis
CC	and treatment of these disorders
SQ	Sequence 2818 BP; 868 A; 548 C; 626 G; 776 T; 0 U; 0 Other;

Query Match 100.0%; Score 87; DB 3; Length 2818;  
 Best Local Similarity 100.0%; Pred. No. 9.5e-37;  
 Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0

0y	1	AGAGAACCCATCATTTAATGAAGTGAATTTTCTGGGCTGAGCTGGACAGAGAGGCAAG	60
Db	1	AGGAAACCCATCATTTAATGAAGTGAATTTTCTGGGCTGAGCTGGACAGAGGCAAG	60
0y	61	AGACACTCTGGACACCACTATGAGACG	87
Db	61	AGACACTCTGGACACCACTATGAGACG	87
RESULT 3			
XX	ID	AAC55313	
XX	AC	AAC55313 standard; DNA; 5514 BP.	
XX	AC	AAC55313;	
XX	DT	05-FEB-2001 (first entry)	
XX	DE	Human activation-induced cytidine deaminase genomic DNA SEQ ID NO:9.	
XX	KM	Activation-induced cytidine deaminase; AID; cytidine deaminase;	
XX	KM	immune related disease; allergy; allergic disease; anti-allergic;	
XX	KM	anti-neuritic; anti-asthmatic; ophthalmological; anti-HIV; dermatological;	
XX	KM	gene therapy; B cell associated immune system disorder; food allergy;	
XX	KM	immunodeficiency disease; immunoglobulin A deficiency disease; asthma;	
XX	KM	IGA nephritis; gamma-globulinemia; atopic dermatitis; allergic colitis;	
XX	KM	drug allergy; allergic rhinitis; Rosen disease; DiGeorge disease; AIDS;	
XX	KM	ataxia telangiectasia; common variable immunodeficiency disorder;	
XX	KM	major histocompatibility class II deficiency disease;	
XX	KM	auto immunodeficiency syndrome; Igg sublaas selection disorder; ds.	
XX	OS	Homo sapiens.	
XX	PN	W0200058480-A1.	
XX	PD	05-OCT-2000.	
XX	PF	28-MAR-2000; 2000MO-UP001918.	
XX	PR	29-MAR-1999; 99GP-00087192.	
XX	PR	24-JUN-1999; 99GP-00178999.	
XX	PR	27-DEC-1999; 99GP-00371382.	
XX	PA	(NIBS ) JAPAN TOBACCO INC.	
XX	PA	(HONT/) HONTO T.	
XX	PI	Honjo T, Muramatsu M;	
XX	DR	WPI; 2000-611715/58.	
XX	PT	Nucleic acid encoding activation induced cytidine deaminase, useful as a	
XX	PT	target for drug development for immune-related diseases including	
XX	PT	allergies.	
XX	PS	Claim 17; Page 142-145; 174pp; Japanese.	
XX	CC	The present invention describes an activation-induced cytidine deaminase	
XX	CC	(AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and has	
XX	CC	cytidine activity similar to APOBEC-1. AID has anti-allergic, anti-neuritic,	
XX	CC	anti-asthmatic, ophthalmological, anti-HIV and dermatological activities,	
XX	CC	and can be used in gene therapy. AID polynucleotides are useful in	
XX	CC	methods for identifying drugs for the treatment of B cell associated	
XX	CC	immune system disorders, immunodeficiency diseases and allergies, such as	
XX	CC	immunoglobulin A (IGA) deficiency disease, IGA nephritis, gamma-	
XX	CC	globulinemia, atopic dermatitis, allergic colitis, asthma, food allergy,	
XX	CC	drug allergy, allergic rhinitis, Rosen disease, DiGeorge disease, ataxia	
XX	CC	telangiectasia, common variable immunodeficiency disorder, MHC (major	
XX	CC	histocompatibility class II deficiency disease, AIDS (auto	
XX	CC	immunodeficiency syndrome), elevated Igg disorder, and Igg sublaas	
XX	CC	selection disorder. The DNA sequences encoding AID may be used for gene	
XX	CC	therapy and the antibodies to the AID protein may be used for diagnosis	
XX	CC	and treatment of these disorders. The present sequence represents a	
XX	CC	genomic DNA sequence of human AID	



SQ Sequence 5514 BP, 1709 A, 1045 C, 1134 G, 1623 T, 0 U, 3 Other;  
Query Match 100.0%; Score 87; DB 3; Length 5514;  
Best Local Similarity 100.0%; Pred. No. 9.6e-37;  
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAGAACCATCATTAATTGAAGTGAAGATTTCCTGGCCCTGAGACTTGACGGAGGCAAGA 60  
DB 1032 AGAGAACCATCATTAATTGAAGTGAAGATTTCCTGGCCCTGAGACTTGACGGAGGCAAGA 1091  
QY 61 AGACACTCTGGACACACCATTAATGACAG 87  
DB 1092 AGACACTCTGGACACACCATTAATGACAG 1118

RESULT 4  
AAC55339  
ID AAC55339 standard; DNA; 11204 BP.  
XX  
XX AAC55339;  
AC  
XX  
DT 05-FEB-2001 (first entry)  
XX  
DE Human activation-induced cytidine deaminase genomic DNA SEQ ID NO:35.  
XX  
XX  
KM Activation-induced cytidine deaminase; AID; cytidine deaminase;  
KM immune related disease; allergy; allergic disease; antiallergic;  
KM antianemic; antisthmatic; ophthalmological; anti-HIV; dermatological;  
KM gene therapy; B cell associated immune system disorder; food allergy;  
KM immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
KM IGA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;  
KM drug allergy; allergic rhinitis; Rosen disease; Digeorge disease; AIDS;  
KM ataxia telangiectasia; common variable immunodeficiency disorder;  
KM major histocompatibility class II deficiency disease;  
KM auto immunodeficiency syndrome; IgG subclass selection disorder; ds.  
XX  
OS Homo sapiens.  
XX  
XX WO200058480-A1.  
XX  
XX  
PD 05-OCT-2000.  
XX  
PF 28-MAR-2000; 2000WO-JP001918.  
XX  
XX  
PR 29-MAR-1999; 99JP-00087192.  
PR 24-JUN-1999; 99JP-00178999.  
PR 27-DEC-1999; 99JP-00371382.  
XX  
XX (NISR ) JAPAN TOBACCO INC.  
XX (HONO/) HONO T.  
XX  
XX Honjo T, Muramatsu M;  
XX  
XX WPI; 2000-611715/58.  
DR  
PT Nucleic acid encoding activation induced cytidine deaminase, useful as a  
PT target for drug development for immune-related diseases including  
PT allergies.  
XX  
XX  
PS Claim 17; Page 163-170; 174pp; Japanese.

The present invention describes an activation-induced cytidine deaminase (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and has cytidine activity similar to APOBEC-1. AID has anti-allergic, antianemic, antiasthmatic, ophthalmological, anti-HIV and dermatological activities, and can be used in gene therapy. AID polynucleotides are useful in methods for identifying drugs for the treatment of B cell associated immune system disorders, immunodeficiency diseases and allergies, such as immunoglobulin A (IGA) deficiency disease, IGA nephritis, gamma-globulinaemia, atopic dermatitis, allergic colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen disease, Digeorge disease, ataxia telangiectasia, common variable immunodeficiency disorder, MHC (major histocompatibility class II deficiency disease, AIDS (auto

CC immunodeficiency syndrome), elevated IgE disorder, and IgG subclass  
CC selection disorder. The DNA sequences encoding AID may be used for gene  
CC therapy and the antibodies to the AID protein may be used for diagnosis  
CC and treatment of these disorders. The present sequence represents a  
CC genomic DNA sequence of human AID  
XX

SQ Sequence 11204 BP, 3305 A, 2273 C, 2373 G, 3253 T, 0 U, 0 Other;  
Query Match 100.0%; Score 87; DB 3; Length 11204;  
Best Local Similarity 100.0%; Pred. No. 9.8e-37;  
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAGAACCATCATTAATTGAAGTGAAGATTTCCTGGCCCTGAGACTTGACGGAGGCAAGA 60  
DB 442 AGAGAACCATCATTAATTGAAGTGAAGATTTCCTGGCCCTGAGACTTGACGGAGGCAAGA 501  
QY 61 AGACACTCTGGACACACCATTAATGACAG 87  
DB 502 AGACACTCTGGACACACCATTAATGACAG 528

RESULT 5  
AB573286  
ID AB573286 standard; DNA; 11204 BP.  
XX  
XX  
AC AB573286;  
XX  
XX  
DT 04-DEC-2002 (first entry)  
XX  
DE DNA encoding human translocation del(12p) protein #1.  
XX  
XX  
XX Chromosome aberration; oncogenic fusion protein; cancer;  
XX proliferative disease; cellular protein isoform; heat shock protein 90;  
KM HSP-90; rheumatoid arthritis; cancer; haematopoietic disorder;  
KM T cell lymphoma; B cell lymphoma; chronic myeloid leukaemia; CML;  
KM acute myeloid leukaemia; AML; chronic myelomonocytic leukaemia; CMML;  
KM papillary thyroid carcinoma; Ewing's sarcoma; melanoma; liposarcoma;  
KM rhabdomyosarcoma; synovial sarcoma; viral infection; gene; ds.  
XX  
XX  
XX Homo sapiens.  
XX  
XX  
XX WO200269900-A2.  
XX  
XX  
PD 12-SEP-2002.  
XX  
PF 01-MAR-2002; 2002WO-US006518.  
XX  
XX  
PR 01-MAR-2001; 2001US-0272751P.  
XX  
XX (CONF-) CONFORMA THERAPEUTICS CORP.  
XX  
XX  
XX Fritz LC, Burrows FJ;  
XX  
XX WPI; 2002-698710/75.  
DR P-PDS; ABG95082.  
XX  
XX  
PT Treating genetically-defined disease associated with chromosomal  
PT aberrations yielding oncogenic fusion proteins, e.g. cell proliferative  
PT diseases, involves administering an inhibitor of heat shock protein 90.  
XX  
XX  
PS Disclosure; Page 242-245; 389pp; English.

The invention describes a method of treating genetically-defined disease associated with chromosomal aberrations yielding oncogenic fusion proteins (II), treating cancerous cells containing (I) in a heterogeneous cell population, treating proliferative diseases associated with mutant protein or cellular protein isoforms (II) dependent on heat shock protein (HSP)-90, or selectively treating cells expressing (II) involving administering HSP90-inhibitor. The method is useful for treating genetically-defined disease with chromosomal aberration yielding oncogenic fusion protein, treating cancerous cells containing fusion protein in heterogeneous cell population, treating proliferative disease



(e.g., rheumatoid arthritis or cancer) associated with mutant protein or cellular protein isoform dependent on heat shock protein (HSP)-90 (e.g., p53), or selectively treating cells expressing mutant protein or cellular protein isoform in a patient heterozygous for (11). The method is useful for treating a disease e.g., hematopoietic disorder such as T or B cell lymphoma, chronic myeloid leukaemia (CML), APL, ALL, AML, NHL and CML, or a disease characterised by a solid tumour such as papillary thyroid carcinoma, Ewing's sarcoma, melanoma, liposarcoma, rhabdomyosarcoma and synovial sarcoma. The method is also useful for treating viral infections. This represents the DNA sequence of a chromosome aberration

Sequence 11204 BP; 3305 A; 2273 C; 2373 G; 3253 T; 0 U; 0 Other;

Query Match 100.0%; Score 87; DB 6; Length 11204;  
Best Local Similarity 100.0%; Pred. No. 9.8e-37;  
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAGAACATCATTAATTAAGTGAATTTTCTGACCTGAGACTTGAGGAGGCAAGA 60  
DB 442 AGAGAACATCATTAATTAAGTGAATTTTCTGACCTGAGACTTGAGGAGGCAAGA 501  
QY 61 AGACACTCTGACACCACTATGACAG 87  
DB 502 AGACACTCTGACACCACTATGACAG 528

RESULT 6  
ABX05468  
ID ABX05468 standard; cDNA; 1543 BP.

AC ABX05468;

DT 17-JAN-2003 (first entry)

XX Human novel polynucleotide #483.

XX Human; gene; sex; genetic disorder; gene mapping; medical imaging; cancer;  
XX neurodegenerative disorder; lymphoid cell disorder; osteoporosis;  
XX Parkinson's disease; Alzheimer's disease; bone degenerative disorder;  
XX osteoarthritis; periodontal disease; liver fibrosis; viral infection;  
XX fungal infection; bacterial infection; autoimmune disease; diabetes;  
XX atopic dermatitis.

XX Homo sapiens.

OS WO200274961-A1.

PN 26-SEP-2002.

XX 14-MAR-2002; 2002WO-US005109.

XX 15-MAR-2001; 2001US-00810173.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Zhou P, Goodrich R, Asundi V, Zhang J, Zhao QA, Ren F,  
XX Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M,  
XX Wehrman T, Wang J, Wang D, Dermanac RT;

XX WPI; 2003-040556/03.

XX P-PSDB; ABU00390.

XX New isolated polypeptides and polynucleotides, useful for preventing,  
XX treating or ameliorating medical conditions, such as cancer,  
XX neurodegenerative disorders, lymphoid cell disorders, bone degenerative  
XX disorders, and infections.

XX Claim 1; SEQ ID NO 483; 235pp; English.

XX The invention relates to human polynucleotides and the polypeptides they  
XX encode. The polynucleotides and polypeptides are useful in diagnostics,  
XX forensics, gene mapping, medical imaging, identification of mutations  
XX responsible for genetic disorders or other traits, assessing biodiversity

CC and producing many other types of data and products dependent on DNA and  
CC amino acid sequences. They are also useful for preventing, treating or  
CC ameliorating medical conditions, such as cancer, neurodegenerative  
CC disorders (e.g., Parkinson's disease, Alzheimer's disease), lymphoid cell  
CC disorders, osteoporosis, osteoarthritis, bone degenerative disorders,  
CC periodontal disease, liver fibrosis, infections (e.g., viral, fungal or  
CC bacterial) or autoimmune diseases (e.g., diabetes, atopic dermatitis).  
CC Sequences ABX04986-ABX0511 represent human polynucleotides of the  
CC invention. Note: The sequence data for this patent is not represented in  
CC the printed specification but is based on sequence information supplied  
CC by the European Patent Office

Sequence 1543 BP; 428 A; 334 C; 352 G; 429 T; 0 U; 0 Other;

Query Match 96.6%; Score 84; DB 7; Length 1543;  
Best Local Similarity 100.0%; Pred. No. 3.8e-35;  
Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GAACCATCATTAATTAAGTGAATTTTCTGACCTGAGACTTGAGGAGGCAAGA 63  
DB 2 GAACCATCATTAATTAAGTGAATTTTCTGACCTGAGACTTGAGGAGGCAAGA 61

QY 64 CACTCTGACACCACTATGACAG 87  
DB 62 CACTCTGACACCACTATGACAG 85

RESULT 7  
ABX73287  
ID ABX73287 standard; DNA; 2791 BP.

AC ABX73287;

DT 04-DEC-2002 (first entry)

XX DNA encoding human translocation del(12p) protein #2.

XX Chromosome aberration; oncogenic fusion protein; cancer;  
XX proliferative disease; cellular protein isoform; heat shock protein 90;  
XX HSP-90; rheumatoid arthritis; cancer; hematopoietic disorder;  
XX T cell lymphoma; B cell lymphoma; chronic myeloid leukaemia; CML;  
XX acute myeloid leukaemia; AML; chronic myelomonocytic leukaemia; CMML;  
XX acute lymphoblastic leukaemia; ALL; APL; NHL; solid tumour;  
XX papillary thyroid carcinoma; Ewing's sarcoma; melanoma; liposarcoma;  
XX rhabdomyosarcoma; synovial sarcoma; viral infection; gene; ds.

XX Homo sapiens.

OS WO200269900-A2.

PN 12-SEP-2002.

XX 01-MAR-2002; 2002WO-US006518.

XX 01-MAR-2001; 2001US-0272751P.

XX (CONF-) CONFORMA THERAPEUTICS CORP.

XX Fritz LC, Burrows FU;

XX WPI; 2002-698710/75.

XX P-PSDB; ABG95083.

XX Treating genetically-defined disease associated with chromosomal  
XX aberrations yielding oncogenic fusion proteins, e.g., cell proliferative  
XX diseases, involves administering an inhibitor of heat shock protein 90.

XX Disclosure; Page 246-247; 389pp; English.

XX The invention describes a method of treating genetically-defined disease  
XX associated with chromosomal aberrations yielding oncogenic fusion  
XX proteins (1) treating cancerous cells containing (1) in a heterogeneous  
XX cell population, treating proliferative diseases associated with mutant



CC protein or cellular protein isoforms (II) dependent on heat shock protein  
CC (HSP)-90, or selectively treating cells expressing (II) involving  
CC administering HSP90-inhibitor. The method is useful for treating  
CC genetically-defined disease with chromosomal aberration yielding  
CC oncogenic fusion protein, treating cancerous cells containing fusion  
CC protein in heterogeneous cell population, treating proliferative disease  
CC (e.g. rheumatoid arthritis or cancer) associated with mutant protein or  
CC cellular protein isoform dependent on heat shock protein (HSP)-90 (e.g.  
CC p53), or selectively treating cells expressing mutant protein or cellular  
CC protein isoform in a patient heterozygous for (II). The method is useful  
CC for treating a disease e.g. haematopoietic disorder such as T or B cell  
CC lymphoma, chronic myeloid leukaemia (CML), APL, ALL, NHL and CML,  
CC or a disease characterised by a solid tumour such as papillary thyroid  
CC carcinoma, Ewing's sarcoma, melanoma, liposarcoma, rhabdomyosarcoma and  
CC synovial sarcoma. The method is also useful for treating viral  
CC infections. This represents the DNA sequence of a chromosome aberration  
XX  
SQ Sequence 2791 BP; 842 A; 548 C; 625 G; 776 T; 0 U; 0 Other;  
Query Match 96.6%; Score 84; DB 6; Length 2791;  
Best Local Similarity 100.0%; Pred. No. 3.9e-35;  
Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 4 GAACCATCATTAATTAAGTAGAGATTCTTCTGGCCTGAGACTTGACGGAGGCAAGAGA 63  
DB 1 GAACCATCATTAATTAAGTAGAGATTCTTCTGGCCTGAGACTTGACGGAGGCAAGAGA 60  
QY 64 CACTCTGGACACCACTATGACAG 87  
DB 61 CACTCTGGACACCACTATGACAG 84  
RESULT 8  
ABS73288  
ID ABS73288 standard; DNA; 2791 BP.  
XX  
AC ABS73288;  
XX  
DT 04-DEC-2002 (first entry)  
XX  
DE DNA encoding human translocation del(12p) protein #3.  
XX  
XX Chromosome aberration; oncogenic fusion protein; cancer;  
KM proliferative disease; cellular protein isoform; heat shock protein 90;  
KM HSP-90; rheumatoid arthritis; cancer; haematopoietic disorder;  
KM T cell lymphoma; B cell lymphoma; chronic myeloid leukaemia; CML;  
KM acute myeloid leukaemia; AML; chronic myelomonocytic leukaemia; CMML;  
KM acute lymphoblastic leukaemia; ALL; APL; NHL; solid tumour;  
KM papillary thyroid carcinoma; Ewing's sarcoma; melanoma; liposarcoma;  
KM rhabdomyosarcoma; synovial sarcoma; viral infection; gene; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200269900-A2.  
XX  
PD 12-SEP-2002.  
XX  
PF 01-MAR-2002; 2002WO-US006518.  
XX  
PR 01-MAR-2001; 2001US-0272751P.  
XX  
PA (CONF-) CONFORMA THERAPEUTICS CORP.  
XX  
PI Fritz LC, Burrows FJ;  
XX  
XX WPI; 2002-698710/75.  
DR P-PSDB; ABG95084.  
XX  
PT Treating genetically-defined disease associated with chromosomal  
PT aberrations yielding oncogenic fusion proteins, e.g. cell proliferative  
PT diseases, involves administering an inhibitor of heat shock protein 90.  
PS Disclosure; Page 246-249; 389pp; English.

XX The invention describes a method of treating genetically-defined disease  
CC associated with chromosomal aberrations yielding oncogenic fusion  
CC proteins (II), treating cancerous cells containing (I) in a heterogeneous  
CC cell population, treating proliferative diseases associated with mutant  
CC protein or cellular protein isoforms (II) dependent on heat shock protein  
CC (HSP)-90, or selectively treating cells expressing (II) involving  
CC administering HSP90-inhibitor. The method is useful for treating  
CC genetically-defined disease with chromosomal aberration yielding  
CC oncogenic fusion protein, treating cancerous cells containing fusion  
CC protein in heterogeneous cell population, treating proliferative disease  
CC (e.g. rheumatoid arthritis or cancer) associated with mutant protein or  
CC cellular protein isoform dependent on heat shock protein (HSP)-90 (e.g.  
CC p53), or selectively treating cells expressing mutant protein or cellular  
CC protein isoform in a patient heterozygous for (II). The method is useful  
CC for treating a disease e.g. haematopoietic disorder such as T or B cell  
CC lymphoma, chronic myeloid leukaemia (CML), APL, ALL, NHL and CML,  
CC or a disease characterised by a solid tumour such as papillary thyroid  
CC carcinoma, Ewing's sarcoma, melanoma, liposarcoma, rhabdomyosarcoma and  
CC synovial sarcoma. The method is also useful for treating viral  
CC infections. This represents the DNA sequence of a chromosome aberration  
XX  
SQ Sequence 2791 BP; 842 A; 548 C; 625 G; 776 T; 0 U; 0 Other;  
Query Match 96.6%; Score 84; DB 6; Length 2791;  
Best Local Similarity 100.0%; Pred. No. 3.9e-35;  
Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 4 GAACCATCATTAATTAAGTAGAGATTCTTCTGGCCTGAGACTTGACGGAGGCAAGAGA 63  
DB 1 GAACCATCATTAATTAAGTAGAGATTCTTCTGGCCTGAGACTTGACGGAGGCAAGAGA 60  
QY 64 CACTCTGGACACCACTATGACAG 87  
DB 61 CACTCTGGACACCACTATGACAG 84  
RESULT 9  
AAC55335  
ID AAC55335 standard; DNA; 26 BP.  
XX  
AC AAC55335;  
XX  
DT 05-FEB-2001 (first entry)  
XX  
DE Human activation-induced cytidine deaminase PCR primer SEQ ID NO:31.  
XX  
XX Activation-induced cytidine deaminase; AID; cytidine deaminase;  
KM immune related disease; allergy; allergic disease; anti-allergic;  
KM anti-naeemic; anti-asthmatic; ophthalmological; anti-HIV; dermatological;  
KM gene therapy; B cell associated immune system disorder; food allergy;  
KM immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
KM IgA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;  
KM drug allergy; allergic rhinitis; Rosen disease; DiGeorge disease; AIDS;  
KM ataxia telangiectasia; common variable immunodeficiency disorder;  
KM major histocompatibility class II deficiency disease; PCR primer;  
KM auto immunodeficiency syndrome; Igg sublaas selection disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200058480-A1.  
XX  
PD 05-OCT-2000.  
XX  
PF 28-MAR-2000; 2000WO-JP001918.  
XX  
XX 29-MAR-1999; 99JP-00087192.  
PR 24-JUN-1999; 99JP-00178999.  
PR 27-DEC-1999; 99JP-00371382.  
XX  
PA (NIBS ) JAPAN TOBACCO INC.  
XX (HONJ/) HONJO T.  
XX



PI Honjo T, Muramatsu M;  
 XX  
 DR MPI; 2000-611715/58.  
 XX  
 PT Nucleic acid encoding activation induced cytidine deaminase, useful as a  
 PT target for drug development for immune-related diseases including  
 PT allergies.  
 XX  
 PS Claim 20; Page 161; 174pp; Japanese.  
 XX  
 CC The present invention describes an activation-induced cytidine deaminase  
 CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and has  
 CC cytidine activity similar to APOBEC-1. AID has anti-allergic, anti-anemic,  
 CC anti-asthmatic, ophthalmological, anti-HIV and dermatological activities,  
 CC and can be used in gene therapy. AID polynucleotides are useful in  
 CC methods for identifying drugs for the treatment of B cell associated  
 CC immune system disorders, immunodeficiency diseases and allergies, such as  
 CC immunoglobulin A (IGA) deficiency disease, IGA nephritis, gamma-  
 CC globulinaemia, atopic dermatitis, allergic colitis, asthma, food allergy,  
 CC drug allergy, allergic rhinitis, rosacea disease, Digeorge disease, ataxia  
 CC telangiectasia, common variable immunodeficiency disorder, MHC (major  
 CC histocompatibility class II deficiency disease, AIDS (auto  
 CC immunodeficiency syndrome), elevated IGE disorder, and IGE subclass  
 CC selection disorder. The DNA sequences encoding AID may be used for gene  
 CC therapy and the antibodies to the AID protein may be used for diagnosis  
 CC and treatment of these disorders. The present sequence represents a  
 CC specifically claimed PCR primer for human AID, which is also used in an  
 CC example from the present invention  
 XX  
 SQ Sequence 26 BP; 7 A; 2 C; 6 G; 11 T; 0 U; 0 Other;  
 QY Query Match 29.9%; Score 26; DB 3; Length 26;  
 Best Local Similarity 100.0%; Pred. No. 0.00049;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 1 CATTAATGAGTGAGATTCTCTG 26  
 QY 11 CATTAATGAGTGAGATTCTCTG 36  
 DB 1 CATTAATGAGTGAGATTCTCTG 26  
 RESULT 10  
 ID ADB88945 standard; DNA; 20 BP.  
 XX  
 AC ADB88945;  
 XX  
 DT 04-DEC-2003 (first entry)  
 XX  
 DE Activation-induced cytidine deaminase mutagenesis method primer #7.  
 XX  
 KM ss; cytosstatic; virucide; antidote; mutation;  
 KM activation-induced cytidine deaminase; cell phenotype;  
 KM monoclonal antibody; antigen; cross-reactivity; tumour;  
 KM germ warfare agent; toxin; V region; primer.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2003061363-A2.  
 XX  
 PD 31-JUL-2003.  
 XX  
 PF 15-JAN-2003; 2003MO-US001149.  
 XX  
 PR 17-JAN-2002; 2002US-0350269P.  
 XX  
 PA (YESH ) UNIV YESHIVA EINSTEIN COLLEGE.  
 XX  
 PI Martin A, Scharff MD;  
 XX  
 DR MPI; 2003-747983/70.  
 XX  
 PT Induction of mutations in gene expressed in eukaryotic cell useful e.g.  
 PT to produce antibodies with greater affinity or specificity for antigens

PT e.g. human pathogens by expressing an activation-induced cytidine  
 PT deaminase gene in cell.  
 XX  
 PS Example 4; Page 34; 73pp; English.  
 XX  
 CC The invention relates to the induction of mutations in a gene expressed  
 CC in a eukaryotic cell, where the gene is operably linked to a promoter and  
 CC within 2 kb of promoter, by expressing a transgenic activation-induced  
 CC cytidine deaminase (AID) gene in the cell. The method is useful for  
 CC producing mutated genes and proteins, determining the effect of a  
 CC mutation on a protein or cell phenotype, and producing genes and proteins  
 CC with altered properties. It is especially useful for producing antibodies  
 CC with altered affinities or specificities for an antigen, or  
 CC increased/decreased cross-reactivity for a second antigen, e.g. to  
 CC produce antibodies with greater affinity/specificity useful  
 CC therapeutically to treat tumours, viruses such as Ebola and Lhaas Fever  
 CC or against germ warfare agents, toxins (e.g. ricin). The method may  
 CC similarly be used to alter affinity, specificity or cross-reactivity of a  
 CC monoclonal antibody. This sequence represents a PCR primer used to  
 CC amplify fragments of the V region of monoclonal antibody genes and the  
 CC AID gene for use in the method of the invention to induce mutations in  
 CC the sequence.  
 XX  
 SQ Sequence 20 BP; 7 A; 4 C; 7 G; 2 T; 0 U; 0 Other;  
 QY Query Match 23.0%; Score 20; DB 8; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 0.82;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 DB 1 GAGGCAAGAAAGACACTCTG 20  
 QY 52 GAGGCAAGAAAGACACTCTG 71  
 DB 1 GAGGCAAGAAAGACACTCTG 20  
 Search completed: March 13, 2004, 05:37:24  
 Job time : 310 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 13, 2004, 05:09:14 ; Search time 60 Seconds  
(without alignments)  
804.679 Million cell updates/sec

Title: US-09-966-880A-11

Perfect score: 87  
Sequence: 1 agagaccatcatatga.....ctgacaccactatgacag 87

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 682709 seqs, 277475446 residues

Word size : 20

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database : Issued Patents NA:\*  
1: /cgn2\_6/prodata/2/ina/5A COMB.seq:\*  
2: /cgn2\_6/prodata/2/ina/5B COMB.seq:\*  
3: /cgn2\_6/prodata/2/ina/5A COMB.seq:\*  
4: /cgn2\_6/prodata/2/ina/5B COMB.seq:\*  
5: /cgn2\_6/prodata/2/ina/PTUS COMB.seq:\*  
6: /cgn2\_6/prodata/2/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
-----					

No matches found

Search completed: March 13, 2004, 09:17:48  
Job time : 61 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 13, 2004, 08:36:41 ; Search time 240 Seconds  
(without alignments)  
1334.579 Million cell updates/sec

Title: US-09-966-880A-11  
Perfect score: 87  
Sequence: 1 agagaccatcatattatgtg.....ctgagaccactatgtacag 87

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 2432557 seqs, 1840798884 residues

Word size : 20

Total number of hits satisfying chosen parameters: 5

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 100 summaries

Database : Published Applications NA:\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
17: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
18: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	87	100.0	87	9	US-09-966-880A-11
2	87	100.0	2818	9	US-09-966-880A-7
3	87	100.0	5514	9	US-09-966-880A-9
4	87	100.0	11204	9	US-09-966-880A-35
5	26	29.9	26	9	US-09-966-880A-31

#### ALIGNMENTS

RESULT 1  
US-09-966-880A-11  
; Sequence 11, Application US/09966880A  
; Patent No. US20020164743A1  
; GENERAL INFORMATION:

APPLICANT: Honjo, Tasuku  
APPLICANT: Muramatsu, Masamichi  
TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
FILE REFERENCE: 06501-088001  
CURRENT APPLICATION NUMBER: US/09/966,880A  
PRIOR FILING DATE: 2001-09-28  
PRIOR APPLICATION NUMBER: PCT/JP00/01918  
PRIOR FILING DATE: 2000-03-28  
PRIOR APPLICATION NUMBER: JP 11-371382  
PRIOR FILING DATE: 1999-12-27  
PRIOR APPLICATION NUMBER: JP 11-178999  
PRIOR FILING DATE: 1999-06-24  
PRIOR APPLICATION NUMBER: JP 11-87192  
PRIOR FILING DATE: 1999-03-29  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 11  
LENGTH: 87  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-966-880A-11

Query Match 100.0%; Score 87; DB 9; Length 87;  
Best local similarity 100.0%; Pred. No. 8.9e-39;  
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGAGACCATCATTTATTTAGTATGATTTTCTGGCTGAGCTTGACGAGGACAGA 60  
DB 1 AGAGACCATCATTTATTTAGTATGATTTTCTGGCTGAGCTTGACGAGGACAGA 60  
DB 61 AGACCTCTGGACACCATTTATTTAGTATGATTTTCTGGCTGAGCTTGACGAGGACAGA 87

RESULT 2  
US-09-966-880A-7  
; Sequence 7, Application US/09966880A  
; Patent No. US20020164743A1  
; GENERAL INFORMATION:  
APPLICANT: Honjo, Tasuku  
APPLICANT: Muramatsu, Masamichi  
TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
FILE REFERENCE: 06501-088001  
CURRENT APPLICATION NUMBER: US/09/966,880A  
PRIOR FILING DATE: 2001-09-28  
PRIOR APPLICATION NUMBER: PCT/JP00/01918  
PRIOR FILING DATE: 2000-03-28  
PRIOR APPLICATION NUMBER: JP 11-371382  
PRIOR FILING DATE: 1999-12-27  
PRIOR APPLICATION NUMBER: JP 11-178999  
PRIOR FILING DATE: 1999-06-24  
PRIOR APPLICATION NUMBER: JP 11-87192  
PRIOR FILING DATE: 1999-03-29  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 7  
LENGTH: 2818  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (80)...(673)  
FEATURE:  
NAME/KEY: 5'UTR  
LOCATION: (1)...(79)  
FEATURE:  
NAME/KEY: 3'UTR  
LOCATION: (677)...(2818)  
US-09-966-880A-7  
Query Match 100.0%; Score 87; DB 9; Length 2818;  
Best local similarity 100.0%; Pred. No. 1.1e-38;







GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 12, 2004, 12:27:59 ; Search time 465.944 Seconds

(without alignments)  
13767.261 Million cell updates/sec

Title: US-09-966-880A-12

Perfect score: 148

Sequence: 1 ccccttgatgaaccgagga.....ttggtatcttcgataag 148

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: 1: gb da: 2: gb hgt: 3: gb in: 4: gb ov: 5: gb pat: 6: gb ph: 7: gb pl: 8: gb pr: 9: gb ro: 10: gb sy: 11: gb un: 12: gb vi: 13: em ba: 14: em fun: 15: em hum: 16: em in: 17: em mu: 18: em om: 19: em or: 20: em ov: 21: em pac: 22: em ph: 23: em pl: 24: em ro: 25: em sy: 26: em un: 27: em vi: 28: em ba: 29: em fun: 30: em hum: 31: em in: 32: em mu: 33: em om: 34: em or: 35: em ov: 36: em pac: 37: em ph: 38: em pl: 39: em ro: 40: em sy: 41: em un: 42: em vi: 43: em ba: 44: em fun: 45: em hum: 46: em in: 47: em mu: 48: em om: 49: em or: 50: em ov: 51: em pac: 52: em ph: 53: em pl: 54: em ro: 55: em sy: 56: em un: 57: em vi: 58: em ba: 59: em fun: 60: em hum: 61: em in: 62: em mu: 63: em om: 64: em or: 65: em ov: 66: em pac: 67: em ph: 68: em pl: 69: em ro: 70: em sy: 71: em un: 72: em vi: 73: em ba: 74: em fun: 75: em hum: 76: em in: 77: em mu: 78: em om: 79: em or: 80: em ov: 81: em pac: 82: em ph: 83: em pl: 84: em ro: 85: em sy: 86: em un: 87: em vi: 88: em ba: 89: em fun: 90: em hum: 91: em in: 92: em mu: 93: em om: 94: em or: 95: em ov: 96: em pac: 97: em ph: 98: em pl: 99: em ro: 100: em sy: 101: em un: 102: em vi: 103: em ba: 104: em fun: 105: em hum: 106: em in: 107: em mu: 108: em om: 109: em or: 110: em ov: 111: em pac: 112: em ph: 113: em pl: 114: em ro: 115: em sy: 116: em un: 117: em vi: 118: em ba: 119: em fun: 120: em hum: 121: em in: 122: em mu: 123: em om: 124: em or: 125: em ov: 126: em pac: 127: em ph: 128: em pl: 129: em ro: 130: em sy: 131: em un: 132: em vi: 133: em ba: 134: em fun: 135: em hum: 136: em in: 137: em mu: 138: em om: 139: em or: 140: em ov: 141: em pac: 142: em ph: 143: em pl: 144: em ro: 145: em sy: 146: em un: 147: em vi: 148: em ba: 149: em fun: 150: em hum: 151: em in: 152: em mu: 153: em om: 154: em or: 155: em ov: 156: em pac: 157: em ph: 158: em pl: 159: em ro: 160: em sy: 161: em un: 162: em vi: 163: em ba: 164: em fun: 165: em hum: 166: em in: 167: em mu: 168: em om: 169: em or: 170: em ov: 171: em pac: 172: em ph: 173: em pl: 174: em ro: 175: em sy: 176: em un: 177: em vi: 178: em ba: 179: em fun: 180: em hum: 181: em in: 182: em mu: 183: em om: 184: em or: 185: em ov: 186: em pac: 187: em ph: 188: em pl: 189: em ro: 190: em sy: 191: em un: 192: em vi: 193: em ba: 194: em fun: 195: em hum: 196: em in: 197: em mu: 198: em om: 199: em or: 200: em ov: 201: em pac: 202: em ph: 203: em pl: 204: em ro: 205: em sy: 206: em un: 207: em vi: 208: em ba: 209: em fun: 210: em hum: 211: em in: 212: em mu: 213: em om: 214: em or: 215: em ov: 216: em pac: 217: em ph: 218: em pl: 219: em ro: 220: em sy: 221: em un: 222: em vi: 223: em ba: 224: em fun: 225: em hum: 226: em in: 227: em mu: 228: em om: 229: em or: 230: em ov: 231: em pac: 232: em ph: 233: em pl: 234: em ro: 235: em sy: 236: em un: 237: em vi: 238: em ba: 239: em fun: 240: em hum: 241: em in: 242: em mu: 243: em om: 244: em or: 245: em ov: 246: em pac: 247: em ph: 248: em pl: 249: em ro: 250: em sy: 251: em un: 252: em vi: 253: em ba: 254: em fun: 255: em hum: 256: em in: 257: em mu: 258: em om: 259: em or: 260: em ov: 261: em pac: 262: em ph: 263: em pl: 264: em ro: 265: em sy: 266: em un: 267: em vi: 268: em ba: 269: em fun: 270: em hum: 271: em in: 272: em mu: 273: em om: 274: em or: 275: em ov: 276: em pac: 277: em ph: 278: em pl: 279: em ro: 280: em sy: 281: em un: 282: em vi: 283: em ba: 284: em fun: 285: em hum: 286: em in: 287: em mu: 288: em om: 289: em or: 290: em ov: 291: em pac: 292: em ph: 293: em pl: 294: em ro: 295: em sy: 296: em un: 297: em vi: 298: em ba: 299: em fun: 300: em hum: 301: em in: 302: em mu: 303: em om: 304: em or: 305: em ov: 306: em pac: 307: em ph: 308: em pl: 309: em ro: 310: em sy: 311: em un: 312: em vi: 313: em ba: 314: em fun: 315: em hum: 316: em in: 317: em mu: 318: em om: 319: em or: 320: em ov: 321: em pac: 322: em ph: 323: em pl: 324: em ro: 325: em sy: 326: em un: 327: em vi: 328: em ba: 329: em fun: 330: em hum: 331: em in: 332: em mu: 333: em om: 334: em or: 335: em ov: 336: em pac: 337: em ph: 338: em pl: 339: em ro: 340: em sy: 341: em un: 342: em vi: 343: em ba: 344: em fun: 345: em hum: 346: em in: 347: em mu: 348: em om: 349: em or: 350: em ov: 351: em pac: 352: em ph: 353: em pl: 354: em ro: 355: em sy: 356: em un: 357: em vi: 358: em ba: 359: em fun: 360: em hum: 361: em in: 362: em mu: 363: em om: 364: em or: 365: em ov: 366: em pac: 367: em ph: 368: em pl: 369: em ro: 370: em sy: 371: em un: 372: em vi: 373: em ba: 374: em fun: 375: em hum: 376: em in: 377: em mu: 378: em om: 379: em or: 380: em ov: 381: em pac: 382: em ph: 383: em pl: 384: em ro: 385: em sy: 386: em un: 387: em vi: 388: em ba: 389: em fun: 390: em hum: 391: em in: 392: em mu: 393: em om: 394: em or: 395: em ov: 396: em pac: 397: em ph: 398: em pl: 399: em ro: 400: em sy: 401: em un: 402: em vi: 403: em ba: 404: em fun: 405: em hum: 406: em in: 407: em mu: 408: em om: 409: em or: 410: em ov: 411: em pac: 412: em ph: 413: em pl: 414: em ro: 415: em sy: 416: em un: 417: em vi: 418: em ba: 419: em fun: 420: em hum: 421: em in: 422: em mu: 423: em om: 424: em or: 425: em ov: 426: em pac: 427: em ph: 428: em pl: 429: em ro: 430: em sy: 431: em un: 432: em vi: 433: em ba: 434: em fun: 435: em hum: 436: em in: 437: em mu: 438: em om: 439: em or: 440: em ov: 441: em pac: 442: em ph: 443: em pl: 444: em ro: 445: em sy: 446: em un: 447: em vi: 448: em ba: 449: em fun: 450: em hum: 451: em in: 452: em mu: 453: em om: 454: em or: 455: em ov: 456: em pac: 457: em ph: 458: em pl: 459: em ro: 460: em sy: 461: em un: 462: em vi: 463: em ba: 464: em fun: 465: em hum: 466: em in: 467: em mu: 468: em om: 469: em or: 470: em ov: 471: em pac: 472: em ph: 473: em pl: 474: em ro: 475: em sy: 476: em un: 477: em vi: 478: em ba: 479: em fun: 480: em hum: 481: em in: 482: em mu: 483: em om: 484: em or: 485: em ov: 486: em pac: 487: em ph: 488: em pl: 489: em ro: 490: em sy: 491: em un: 492: em vi: 493: em ba: 494: em fun: 495: em hum: 496: em in: 497: em mu: 498: em om: 499: em or: 500: em ov: 501: em pac: 502: em ph: 503: em pl: 504: em ro: 505: em sy: 506: em un: 507: em vi: 508: em ba: 509: em fun: 510: em hum: 511: em in: 512: em mu: 513: em om: 514: em or: 515: em ov: 516: em pac: 517: em ph: 518: em pl: 519: em ro: 520: em sy: 521: em un: 522: em vi: 523: em ba: 524: em fun: 525: em hum: 526: em in: 527: em mu: 528: em om: 529: em or: 530: em ov: 531: em pac: 532: em ph: 533: em pl: 534: em ro: 535: em sy: 536: em un: 537: em vi: 538: em ba: 539: em fun: 540: em hum: 541: em in: 542: em mu: 543: em om: 544: em or: 545: em ov: 546: em pac: 547: em ph: 548: em pl: 549: em ro: 550: em sy: 551: em un: 552: em vi: 553: em ba: 554: em fun: 555: em hum: 556: em in: 557: em mu: 558: em om: 559: em or: 560: em ov: 561: em pac: 562: em ph: 563: em pl: 564: em ro: 565: em sy: 566: em un: 567: em vi: 568: em ba: 569: em fun: 570: em hum: 571: em in: 572: em mu: 573: em om: 574: em or: 575: em ov: 576: em pac: 577: em ph: 578: em pl: 579: em ro: 580: em sy: 581: em un: 582: em vi: 583: em ba: 584: em fun: 585: em hum: 586: em in: 587: em mu: 588: em om: 589: em or: 590: em ov: 591: em pac: 592: em ph: 593: em pl: 594: em ro: 595: em sy: 596: em un: 597: em vi: 598: em ba: 599: em fun: 600: em hum: 601: em in: 602: em mu: 603: em om: 604: em or: 605: em ov: 606: em pac: 607: em ph: 608: em pl: 609: em ro: 610: em sy: 611: em un: 612: em vi: 613: em ba: 614: em fun: 615: em hum: 616: em in: 617: em mu: 618: em om: 619: em or: 620: em ov: 621: em pac: 622: em ph: 623: em pl: 624: em ro: 625: em sy: 626: em un: 627: em vi: 628: em ba: 629: em fun: 630: em hum: 631: em in: 632: em mu: 633: em om: 634: em or: 635: em ov: 636: em pac: 637: em ph: 638: em pl: 639: em ro: 640: em sy: 641: em un: 642: em vi: 643: em ba: 644: em fun: 645: em hum: 646: em in: 647: em mu: 648: em om: 649: em or: 650: em ov: 651: em pac: 652: em ph: 653: em pl: 654: em ro: 655: em sy: 656: em un: 657: em vi: 658: em ba: 659: em fun: 660: em hum: 661: em in: 662: em mu: 663: em om: 664: em or: 665: em ov: 666: em pac: 667: em ph: 668: em pl: 669: em ro: 670: em sy: 671: em un: 672: em vi: 673: em ba: 674: em fun: 675: em hum: 676: em in: 677: em mu: 678: em om: 679: em or: 680: em ov: 681: em pac: 682: em ph: 683: em pl: 684: em ro: 685: em sy: 686: em un: 687: em vi: 688: em ba: 689: em fun: 690: em hum: 691: em in: 692: em mu: 693: em om: 694: em or: 695: em ov: 696: em pac: 697: em ph: 698: em pl: 699: em ro: 700: em sy: 701: em un: 702: em vi: 703: em ba: 704: em fun: 705: em hum: 706: em in: 707: em mu: 708: em om: 709: em or: 710: em ov: 711: em pac: 712: em ph: 713: em pl: 714: em ro: 715: em sy: 716: em un: 717: em vi: 718: em ba: 719: em fun: 720: em hum: 721: em in: 722: em mu: 723: em om: 724: em or: 725: em ov: 726: em pac: 727: em ph: 728: em pl: 729: em ro: 730: em sy: 731: em un: 732: em vi: 733: em ba: 734: em fun: 735: em hum: 736: em in: 737: em mu: 738: em om: 739: em or: 740: em ov: 741: em pac: 742: em ph: 743: em pl: 744: em ro: 745: em sy: 746: em un: 747: em vi: 748: em ba: 749: em fun: 750: em hum: 751: em in: 752: em mu: 753: em om: 754: em or: 755: em ov: 756: em pac: 757: em ph: 758: em pl: 759: em ro: 760: em sy: 761: em un: 762: em vi: 763: em ba: 764: em fun: 765: em hum: 766: em in: 767: em mu: 768: em om: 769: em or: 770: em ov: 771: em pac: 772: em ph: 773: em pl: 774: em ro: 775: em sy: 776: em un: 777: em vi: 778: em ba: 779: em fun: 780: em hum: 781: em in: 782: em mu: 783: em om: 784: em or: 785: em ov: 786: em pac: 787: em ph: 788: em pl: 789: em ro: 790: em sy: 791: em un: 792: em vi: 793: em ba: 794: em fun: 795: em hum: 796: em in: 797: em mu: 798: em om: 799: em or: 800: em ov: 801: em pac: 802: em ph: 803: em pl: 804: em ro: 805: em sy: 806: em un: 807: em vi: 808: em ba: 809: em fun: 810: em hum: 811: em in: 812: em mu: 813: em om: 814: em or: 815: em ov: 816: em pac: 817: em ph: 818: em pl: 819: em ro: 820: em sy: 821: em un: 822: em vi: 823: em ba: 824: em fun: 825: em hum: 826: em in: 827: em mu: 828: em om: 829: em or: 830: em ov: 831: em pac: 832: em ph: 833: em pl: 834: em ro: 835: em sy: 836: em un: 837: em vi: 838: em ba: 839: em fun: 840: em hum: 841: em in: 842: em mu: 843: em om: 844: em or: 845: em ov: 846: em pac: 847: em ph: 848: em pl: 849: em ro: 850: em sy: 851: em un: 852: em vi: 853: em ba: 854: em fun: 855: em hum: 856: em in: 857: em mu: 858: em om: 859: em or: 860: em ov: 861: em pac: 862: em ph: 863: em pl: 864: em ro: 865: em sy: 866: em un: 867: em vi: 868: em ba: 869: em fun: 870: em hum: 871: em in: 872: em mu: 873: em om: 874: em or: 875: em ov: 876: em pac: 877: em ph: 878: em pl: 879: em ro: 880: em sy: 881: em un: 882: em vi: 883: em ba: 884: em fun: 885: em hum: 886: em in: 887: em mu: 888: em om: 889: em or: 890: em ov: 891: em pac: 892: em ph: 893: em pl: 894: em ro: 895: em sy: 896: em un: 897: em vi: 898: em ba: 899: em fun: 900: em hum: 901: em in: 902: em mu: 903: em om: 904: em or: 905: em ov: 906: em pac: 907: em ph: 908: em pl: 909: em ro: 910: em sy: 911: em un: 912: em vi: 913: em ba: 914: em fun: 915: em hum: 916: em in: 917: em mu: 918: em om: 919: em or: 920: em ov: 921: em pac: 922: em ph: 923: em pl: 924: em ro: 925: em sy: 926: em un: 927: em vi: 928: em ba: 929: em fun: 930: em hum: 931: em in: 932: em mu: 933: em om: 934: em or: 935: em ov: 936: em pac: 937: em ph: 938: em pl: 939: em ro: 940: em sy: 941: em un: 942: em vi: 943: em ba: 944: em fun: 945: em hum: 946: em in: 947: em mu: 948: em om: 949: em or: 950: em ov: 951: em pac: 952: em ph: 953: em pl: 954: em ro: 955: em sy: 956: em un: 957: em vi: 958: em ba: 959: em fun: 960: em hum: 961: em in: 962: em mu: 963: em om: 964: em or: 965: em ov: 966: em pac: 967: em ph: 968: em pl: 969: em ro: 970: em sy: 971: em un: 972: em vi: 973: em ba: 974: em fun: 975: em hum: 976: em in: 977: em mu: 978: em om: 979: em or: 980: em ov: 981: em pac: 982: em ph: 983: em pl: 984: em ro: 985: em sy: 986: em un: 987: em vi: 988: em ba: 989: em fun: 990: em hum: 991: em in: 992: em mu: 993: em om: 994: em or: 995: em ov: 996: em pac: 997: em ph: 998: em pl: 999: em ro: 1000: em sy: 1001: em un: 1002: em vi: 1003: em ba: 1004: em fun: 1005: em hum: 1006: em in: 1007: em mu: 1008: em om: 1009: em or: 1010: em ov: 1011: em pac: 1012: em ph: 1013: em pl: 1014: em ro: 1015: em sy: 1016: em un: 1017: em vi: 1018: em ba: 1019: em fun: 1020: em hum: 1021: em in: 1022: em mu: 1023: em om: 1024: em or: 1025: em ov: 1026: em pac: 1027: em ph: 1028: em pl: 1029: em ro: 1030: em sy: 1031: em un: 1032: em vi: 1033: em ba: 1034: em fun: 1035: em hum: 1036: em in: 1037: em mu: 1038: em om: 1039: em or: 1040: em ov: 1041: em pac: 1042: em ph: 1043: em pl: 1044: em ro: 1045: em sy: 1046: em un: 1047: em vi: 1048: em ba: 1049: em fun: 1050: em hum: 1051: em in: 1052: em mu: 1053: em om: 1054: em or: 1055: em ov: 1056: em pac: 1057: em ph: 1058: em pl: 1059: em ro: 1060: em sy: 1061: em un: 1062: em vi: 1063: em ba: 1064: em fun: 1065: em hum: 1066: em in: 1067: em mu: 1068: em om: 1069: em or: 1070: em ov: 1071: em pac: 1072: em ph: 1073: em pl: 1074: em ro: 1075: em sy: 1076: em un: 1077: em vi: 1078: em ba: 1079: em fun: 1080: em hum: 1081: em in: 1082: em mu: 1083: em om: 1084: em or: 1085: em ov: 1086: em pac: 1087: em ph: 1088: em pl: 1089: em ro: 1090: em sy: 1091: em un: 1092: em vi: 1093: em ba: 1094: em fun: 1095: em hum: 1096: em in: 1097: em mu: 1098: em om: 1099: em or: 1100: em ov: 1101: em pac: 1102: em ph: 1103: em pl: 1104: em ro: 1105: em sy: 1106: em un: 1107: em vi: 1108: em ba: 1109: em fun: 1110: em hum: 1111: em in: 1112: em mu: 1113: em om: 1114: em or: 1115: em ov: 1116: em pac: 1117: em ph: 1118: em pl: 1119: em ro: 1120: em sy: 1121: em un: 1122: em vi: 1123: em ba: 1124: em fun: 1125: em hum: 1126: em in: 1127: em mu: 1128: em om: 1129: em or: 1130: em ov: 1131: em pac: 1132: em ph: 1133: em pl: 1134: em ro: 1135: em sy: 1136: em un: 1137: em vi: 1138: em ba: 1139: em fun: 1140: em hum: 1141: em in: 1142: em mu: 1143: em om: 1144: em or: 1145: em ov: 1146: em pac: 1147: em ph: 1148: em pl: 1149: em ro: 1150: em sy: 1151: em un: 1152: em vi: 1153: em ba: 1154: em fun: 1155: em hum: 1156: em in: 1157: em mu: 1158: em om: 1159: em or: 1160: em ov: 1161: em pac: 1162: em ph: 1163: em pl: 1164: em ro: 1165: em sy: 1166: em un: 1167: em vi: 1168: em ba: 1169: em fun: 1170: em hum: 1171: em in: 1172: em mu: 1173: em om: 1174: em or: 1175: em ov: 1176: em pac: 1177: em ph: 1178: em pl: 1179: em ro: 1180: em sy: 1181: em un: 1182: em vi: 1183: em ba: 1184: em fun: 1185: em hum: 1186: em in: 1187: em mu: 1188: em om: 1189: em or: 1190: em ov: 1191: em pac: 1192: em ph: 1193: em pl: 1194: em ro: 1195: em sy: 1196: em un: 1197: em vi: 1198: em ba: 1199: em fun: 1200: em hum: 1201: em in: 1202: em mu: 1203: em om: 1204: em or: 1205: em ov: 1206: em pac: 1207: em ph: 1208: em pl: 1209: em ro: 1210: em sy: 1211: em un: 1212: em vi: 1213: em ba: 1214: em fun: 1215: em hum: 1216: em in: 1217: em mu: 1218: em om: 1219: em or: 1220: em ov: 1221: em pac: 1222: em ph: 1223: em pl: 1224: em ro: 1225: em sy: 1226: em un: 1227: em vi: 1228: em ba: 1229: em fun: 1230: em hum: 1231: em in: 1232: em mu: 1233: em om: 1234: em or: 1235: em ov: 1236: em pac: 1237: em ph: 1238: em pl: 1239: em ro: 1240: em sy: 1241: em un: 1242: em vi: 1243: em ba: 1244: em fun: 1245: em hum: 1246: em in: 1247: em mu: 1248: em om: 1249: em or: 1250: em ov: 1251: em pac: 1252: em ph: 1253: em pl: 1254: em ro: 1255: em sy: 1256: em un: 1257: em vi: 1258: em ba: 1259: em fun: 1260: em hum: 1261: em in: 1262: em mu: 1263: em om: 1264: em or: 1265: em ov: 1266: em pac: 1267: em ph: 1268: em pl: 1269: em ro: 1270: em sy: 1271: em un: 1272: em vi: 1273: em ba: 1274: em fun: 1275: em hum: 1276: em in: 1277: em mu: 1278: em om: 1279: em or: 1280: em ov: 1281: em pac: 1282: em ph: 1283: em pl: 1284: em ro: 1285: em sy: 1286: em un: 1287: em vi: 1288: em ba: 1289: em fun: 1290: em hum: 1291: em in: 1292: em mu: 1293: em om: 1294: em or: 1295: em ov: 1296: em pac: 1297: em ph: 1298: em pl: 1299: em ro: 1300: em sy: 1301: em un: 1302: em vi: 1303: em ba: 1304: em fun: 1305: em hum: 1306: em in: 1307: em mu: 1308: em om: 1309: em or: 1310: em ov: 1311: em pac: 1312: em ph: 1313: em pl: 1314: em ro: 1315: em sy: 1316: em un: 1317: em vi: 1318: em ba: 1319: em fun: 1320: em hum: 1321: em in: 1322: em mu: 1323: em om: 1324: em or: 1325: em ov: 1326: em pac: 1327: em ph: 1328: em pl: 1329: em ro: 1330: em sy: 1331: em un: 1332: em vi: 1333: em ba: 1334: em fun: 1335: em hum: 1336: em in: 1337: em mu: 1338: em om: 1339: em or: 1340: em ov: 1341: em pac: 1342: em ph: 1343: em pl: 1344: em ro: 1345: em sy: 1346: em un: 1347: em vi: 1348: em ba: 1349: em fun: 1350: em hum: 1351: em in: 1352: em mu: 1353: em om: 1354: em or: 1355: em ov: 1356: em pac: 1357: em ph: 1358: em pl: 1359: em ro: 1360: em sy: 1361: em un: 1362: em vi: 1363: em ba: 1364: em fun: 1365: em hum: 1366: em in: 1367: em mu: 1368: em om: 1369: em or: 1370: em ov: 1371: em pac: 1372: em ph: 1373: em pl: 1374: em ro: 1375: em sy: 1376: em un: 1377: em vi: 1378: em ba: 1379: em fun: 1380: em hum: 1381: em in: 1382: em mu: 1383: em om: 1384: em or: 1385: em ov: 1386: em pac: 1387: em ph: 1388: em pl: 1389: em ro: 1390: em sy: 1391: em un: 1392: em vi: 13



COMMENT JAPAN TOBACCO INC, TASUKU HONJO  
OS Homo sapiens (human)  
PN JP 2001245669-A/10  
PD 11-SEP-2001  
PF 28-MAR-2000 JP 2000092981  
PI TASUKU HONJO, MASAMICHI MURAMATSU  
PC C12N15/09, A61K39/395, A61P1/00, A61P11/06, A61P13/12,  
PC A61P17/00,  
PC A61P27/02, A61P27/16, A61P37/02, A61P37/08, C07K16/18, C12N1/19, PC  
C12N1/21,  
PC C12N5/10, C12N9/78, C12P21/02, C12P21/08, C12N1/21, C12P1/19, PC  
(C12N5/10, C12P1/19), C12N15/00, C12N5/00, C12N5/00, C12N1/91) CC  
FH Key Location/Qualifiers  
FEATURES  
source 1. 148  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Query Match 100.0%; Score 148; DB 6; Length 148;  
Best Local Similarity 100.0%; Pred. No. 1.5e-40;  
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CCTCTGATGAAACCGAGAGAGTTCTTTTACCAATTCMAAAATGTCGGCTGAGGCTAAGG 60  
DB 1 CCTCTGATGAAACCGAGAGAGTTCTTTTACCAATTCMAAAATGTCGGCTGAGGCTAAGG 60  
QY 61 TCGGCGTGAACCTACCTGCTGCTACGTAAGAGAGCGTGAAGAGCTGCTACATCTTTTC 120  
DB 61 TCGGCGTGAACCTACCTGCTGCTACGTAAGAGAGCGTGAAGAGCTGCTACATCTTTTC 120  
QY 121 ACTGACTTGGTATCTTCGCAATTAAG 148  
DB 121 ACTGACTTGGTATCTTCGCAATTAAG 148  
RESULT 2  
AF529834 595 bp mRNA linear SYN 17-SEP-2002  
LOCUS AF529834  
DEFINITION Mus musculus clone 7 transgenic Homo sapiens AID (AID) mRNA,  
complete cds.  
ACCESSION AF529834  
VERSION AF529834.1 GI:22297255  
KEYWORDS Mus musculus (house mouse)  
SOURCE Mus musculus  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 595)  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
AUTHORS Martin, A. and Scharff, M.D.  
TITLE Somatic hypermutation of the AID transgene in B and non-B cells  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12304-12308 (2002)  
REFERENCE 2 (bases 1 to 595)  
Martin, A. and Scharff, M.D.  
AUTHORS Martin, A. and Scharff, M.D.  
TITLE Direct Submission  
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of  
Medicine, 1300 Morris Park Ave, Chanin 404, Bronx, NY 10461, USA  
FEATURES  
source 1. 595  
Location/Qualifiers  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="7"  
/cell\_type="hybridoma Pl-5"  
/transgenic  
1. 595  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
1. 595  
/gene="AID"  
1. 595  
CDS  
1. 595  
/gene="AID"  
1. 595  
CDS  
1. 595

/gene="AID"  
/codon\_start=1  
/product="AID"  
/protein\_id="AA05421.1"  
/db\_xref="GI:22297256"  
/translation="MDSILMMRRKFLYQFNVMNAKGRRETYLCYVVKRDSATSPSL  
DFGYLRNKGCHVELFLRYISDMDLDPGRCYVMTFTSPSPCYDCARHVAADFLRGNP  
NLSRIPTALRYFCEDRKAEFGRLRLHRAVGQIAIWTFFDYFCWNIFVENHRTFK  
AMEGLHNSVRLSRQLRILLPLYEVDDLDAPFTLGL"  
ORIGIN  
Query Match 100.0%; Score 148; DB 12; Length 595;  
Best Local Similarity 100.0%; Pred. No. 1.7e-40;  
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTCTGATGAAACCGAGAGAGTTCTTTTACCAATTCMAAAATGTCGGCTGAGGCTAAGG 60  
DB 9 CCTCTGATGAAACCGAGAGAGTTCTTTTACCAATTCMAAAATGTCGGCTGAGGCTAAGG 68  
QY 61 TCGGCGTGAACCTACCTGCTGCTACGTAAGAGAGCGTGAAGAGCTGCTACATCTTTTC 120  
DB 69 TCGGCGTGAACCTACCTGCTGCTACGTAAGAGAGCGTGAAGAGCTGCTACATCTTTTC 128  
QY 121 ACTGACTTGGTATCTTCGCAATTAAG 148  
DB 129 ACTGACTTGGTATCTTCGCAATTAAG 156  
RESULT 3  
AF529815 596 bp mRNA linear PRI 17-SEP-2002  
LOCUS AF529815  
DEFINITION Homo sapiens clone Ramos 1 AID (AID) mRNA, partial cds.  
ACCESSION AF529815  
VERSION AF529815.1 GI:22297217  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 596)  
Martin, A. and Scharff, M.D.  
TITLE Somatic hypermutation of the AID transgene in B and non-B cells  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12304-12308 (2002)  
REFERENCE 2 (bases 1 to 596)  
Martin, A. and Scharff, M.D.  
AUTHORS Martin, A. and Scharff, M.D.  
TITLE Direct Submission  
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of  
Medicine, 1300 Morris Park Ave, Chanin 404, Bronx, NY 10461, USA  
FEATURES  
source 1. 596  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="Ramos 1"  
1. 596  
/gene="AID"  
1. 596  
/note="Integrated into Burkitt's lymphoma cell line Ramos"  
/codon\_start=1  
/product="AID"  
/protein\_id="AA05402.1"  
/db\_xref="GI:22297218"  
/translation="MDSILMMRRKFLYQFNVMNAKGRRETYLCYVVKRDSATSPSL  
DFGYLRNKGCHVELFLRYISDMDLDPGRCYVMTFTSPSPCYDCARHVAADFLRGNP  
NLSRIPTALRYFCEDRKAEFGRLRLHRAVGQIAIWTFFDYFCWNIFVENHRTFK  
AMEGLHNSVRLSRQLRILLPLYEVDDLDAPFTLGL"  
ORIGIN  
Query Match 100.0%; Score 148; DB 9; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.7e-40;  
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



QY 1 CCTCTGATGACCGGAGGAAGTTCTTTTACCAATTCATAAATGTCCTGCGCTAAGG 60  
Db 9 CCTCTGATGACCGGAGGAAGTTCTTTTACCAATTCATAAATGTCCTGCGCTAAGG 68  
QY 61 TCGGCGGAGACCTACCTGCTACGTAAGAGGCGGTGACAGTGCATACCTTTTC 120  
Db 69 TCGGCGGAGACCTACCTGCTACGTAAGAGGCGGTGACAGTGCATACCTTTTC 128  
QY 121 ACTGACTTGGTTATCTTCGCAATTAAG 148  
Db 129 ACTGACTTGGTTATCTTCGCAATTAAG 156

RESULT 4  
AF529816 596 bp mRNA linear PRI 17-SEP-2002  
LOCUS Homo sapiens clone Ramos 2 AID (AID) mRNA, partial cds.  
DEFINITION AF529816  
ACCESSION AF529816 GI:22297219  
VERSION  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM  
REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
TITLE 1 (bases 1 to 596)  
JOURNAL Martin, A. and Scharff, M.D.  
AUTHORS Somatic hypermutation of the AID transgene in B and non-B cells  
PROC. Natl. Acad. Sci. U.S.A. 99 (19), 12304-12308 (2002)  
REFERENCE 2 (bases 1 to 596)  
AUTHORS Martin, A. and Scharff, M.D.  
TITLE Direct Submission  
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of  
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA

FEATURES  
source  
1. 596  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="Ramos 2"  
1. .>596  
/gene="AID"  
1. .>596  
/note="Integrated into Burkitt's lymphoma cell line Ramos"  
/codon\_start=1  
/product="AID"  
/protein\_id="AA95403.1"  
/db\_xref="GI:22297220"  
/translation="MDSLIMNRKPLVOPKVMKGRRETYLCVYKRDATSTSL  
DFGLRNKNGCHVELFLKRTISDMDLPGRCHRVYVTFSTSPCYDCARVADLRGNP  
NLSLRIFLARLYFCEDRKAPBEGRLRLRAGVQIALMTFQDPCVWNTVEVHERTEK  
AMEGLHENSVALSRQLRRLILPLVEVDLDRDAFRTLGL"

ORIGIN  
Query Match 100.0%; Score 148; DB 9; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.7e-40;  
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTCTGATGACCGGAGGAAGTTCTTTTACCAATTCATAAATGTCCTGCGCTAAGG 60  
Db 9 CCTCTGATGACCGGAGGAAGTTCTTTTACCAATTCATAAATGTCCTGCGCTAAGG 68  
QY 61 TCGGCGGAGACCTACCTGCTACGTAAGAGGCGGTGACAGTGCATACCTTTTC 120  
Db 69 TCGGCGGAGACCTACCTGCTACGTAAGAGGCGGTGACAGTGCATACCTTTTC 128  
QY 121 ACTGACTTGGTTATCTTCGCAATTAAG 148  
Db 129 ACTGACTTGGTTATCTTCGCAATTAAG 156

RESULT 5  
AF529817

LOCUS AF529817 596 bp mRNA linear PRI 17-SEP-2002  
DEFINITION Homo sapiens clone Ramos 3 AID (AID) mRNA, partial cds.  
ACCESSION AF529817  
VERSION AF529817.1 GI:22297221  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM  
REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
TITLE 1 (bases 1 to 596)  
JOURNAL Martin, A. and Scharff, M.D.  
AUTHORS Somatic hypermutation of the AID transgene in B and non-B cells  
PROC. Natl. Acad. Sci. U.S.A. 99 (19), 12304-12308 (2002)  
REFERENCE 2 (bases 1 to 596)  
AUTHORS Martin, A. and Scharff, M.D.  
TITLE Direct Submission  
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of  
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA

FEATURES  
source  
1. 596  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="Ramos 3"  
1. .>596  
/gene="AID"  
1. .>596  
/note="Integrated into Burkitt's lymphoma cell line Ramos"  
/codon\_start=1  
/product="AID"  
/protein\_id="AA95404.1"  
/db\_xref="GI:22297222"  
/translation="MDSLIMNRKPLVOPKVMKGRRETYLCVYKRDATSTSL  
DFGLRNKNGCHVELFLKRTISDMDLPGRCHRVYVTFSTSPCYDCARVADLRGNP  
NLSLRIFLARLYFCEDRKAPBEGRLRLRAGVQIALMTFQDPCVWNTVEVHERTEK  
AMEGLHENSVALSRQLRRLILPLVEVDLDRDAFRTLGL"

ORIGIN  
Query Match 100.0%; Score 148; DB 9; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.7e-40;  
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTCTGATGACCGGAGGAAGTTCTTTTACCAATTCATAAATGTCCTGCGCTAAGG 60  
Db 9 CCTCTGATGACCGGAGGAAGTTCTTTTACCAATTCATAAATGTCCTGCGCTAAGG 68  
QY 61 TCGGCGGAGACCTACCTGCTACGTAAGAGGCGGTGACAGTGCATACCTTTTC 120  
Db 69 TCGGCGGAGACCTACCTGCTACGTAAGAGGCGGTGACAGTGCATACCTTTTC 128  
QY 121 ACTGACTTGGTTATCTTCGCAATTAAG 148  
Db 129 ACTGACTTGGTTATCTTCGCAATTAAG 156

RESULT 6  
AF529818 596 bp mRNA linear PRI 17-SEP-2002  
LOCUS Homo sapiens clone Ramos 4 truncated AID (AID) mRNA, complete cds.  
DEFINITION AF529818  
ACCESSION AF529818 GI:22297223  
VERSION  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM  
REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
TITLE 1 (bases 1 to 596)  
JOURNAL Martin, A. and Scharff, M.D.  
AUTHORS Somatic hypermutation of the AID transgene in B and non-B cells  
PROC. Natl. Acad. Sci. U.S.A. 99 (19), 12304-12308 (2002)  
REFERENCE 2 (bases 1 to 596)  
AUTHORS Martin, A. and Scharff, M.D.



TITLE Direct Submission  
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA  
FEATURES  
source  
1. 596  
/organism="Homo sapiens"  
/mol\_type="rRNA"  
/db\_xref="taxon:9606"  
/clone="Ramos 4"  
1. 596  
/gene="AID"  
1. 570  
/note="Integrated into Burkitt's lymphoma cell line Ramos"  
/codon\_start=1  
/product="truncated AID"  
/protein\_id="AA095405.1"  
/db\_xref="GI:22297224"  
/translation="MDSLIMNRKFLYOPKQVWAKGRRETYLCYVVKRDSATSPSL  
DFGLRNKNGCHVELLFLRYISDMDDPGCVRVTFWTSPPCYDCARHVDPLRGNP  
NLSIRIFPARLYFCEDHKAPBEGRLRHAGVQIALMTPEKDYFCMNTFVENHERTFK  
AMEGLHENSRLSRQLRIILPLVEVDLJL"

ORIGIN  
Query Match 100.0%; Score 148; DB 9; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.7e-40;  
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTCTGATGACCGGAGGAAGTTCTTTACCAATTCAAAATGTCGGCTGAGG 60  
DB 9 CCTCTGATGACCGGAGGAAGTTCTTTACCAATTCAAAATGTCGGCTGAGG 68  
QY 61 TCGGCGTGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120  
DB 69 TCGGCGTGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 128  
QY 121 ACTGACTTGGTTATCTTCCCAATTAAG 148  
DB 129 ACTGACTTGGTTATCTTCCCAATTAAG 156

RESULT 7  
AF529819 596 bp mRNA linear PRI 17-SEP-2002  
LOCUS Homo sapiens clone Ramos 5 AID (AID) mRNA, partial cds.  
ACCESSION AF529819  
VERSION AF529819.1 GI:22297225  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 596)  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 596)  
Somatic hypermutation of the AID transgene in B and non-B cells  
Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12304-12308 (2002)  
2 (bases 1 to 596)  
Martin, A. and Scharff, M.D.  
Direct Submission  
Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of  
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA  
JOURNAL Location/Qualifiers  
1. 596  
/organism="Homo sapiens"  
/mol\_type="rRNA"  
/db\_xref="taxon:9606"  
/clone="Ramos 5"  
1. 596  
/gene="AID"  
1. 596  
/note="Integrated into Burkitt's lymphoma cell line Ramos"  
/codon\_start=1

FEATURES  
source  
1. 596  
/organism="Homo sapiens"  
/mol\_type="rRNA"  
/db\_xref="taxon:9606"  
/clone="Ramos 5"  
1. 596  
/gene="AID"  
1. 596  
/note="Integrated into Burkitt's lymphoma cell line Ramos"  
/codon\_start=1

/product="AID"  
/protein\_id="AA095406.1"  
/db\_xref="GI:22297226"  
/translation="MDSLIMNRKFLYOPKQVWAKGRRETYLCYVVKRDSATSPSL  
DFGLRNKNGCHVELLFLRYISDMDDPGCVRVTFWTSPPCYDCARHVDPLRGNP  
NLSIRIFPARLYFCEDHKAPBEGRLRHAGVQIALMTPEKDYFCMNTFVENHERTFK  
AMEGLHENSRLSRQLRIILPLVEVDLJL"

ORIGIN  
Query Match 100.0%; Score 148; DB 9; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.7e-40;  
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTCTGATGACCGGAGGAAGTTCTTTACCAATTCAAAATGTCGGCTGAGG 60  
DB 9 CCTCTGATGACCGGAGGAAGTTCTTTACCAATTCAAAATGTCGGCTGAGG 68  
QY 61 TCGGCGTGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120  
DB 69 TCGGCGTGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 128  
QY 121 ACTGACTTGGTTATCTTCCCAATTAAG 148  
DB 129 ACTGACTTGGTTATCTTCCCAATTAAG 156

RESULT 8  
AF529820 596 bp mRNA linear PRI 17-SEP-2002  
LOCUS Homo sapiens clone Ramos 6 AID (AID) mRNA, partial cds.  
ACCESSION AF529820  
VERSION AF529820.1 GI:22297227  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 596)  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 596)  
Somatic hypermutation of the AID transgene in B and non-B cells  
Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12304-12308 (2002)  
2 (bases 1 to 596)  
Martin, A. and Scharff, M.D.  
Direct Submission  
Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of  
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA  
JOURNAL Location/Qualifiers  
1. 596  
/organism="Homo sapiens"  
/mol\_type="rRNA"  
/db\_xref="taxon:9606"  
/clone="Ramos 6"  
1. 596  
/gene="AID"  
1. 596  
/note="Integrated into Burkitt's lymphoma cell line Ramos"  
/codon\_start=1  
/product="AID"  
/protein\_id="AA095407.1"  
/db\_xref="GI:22297228"  
/translation="MDSLIMNRKFLYOPKQVWAKGRRETYLCYVVKRDSATSPSL  
DFGLRNKNGCHVELLFLRYISDMDDPGCVRVTFWTSPPCYDCARHVDPLRGNP  
NLSIRIFPARLYFCEDHKAPBEGRLRHAGVQIALMTPEKDYFCMNTFVENHERTFK  
AMEGLHENSRLSRQLRIILPLVEVDLJL"

ORIGIN  
Query Match 100.0%; Score 148; DB 9; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.7e-40;  
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTCTGATGACCGGAGGAAGTTCTTTACCAATTCAAAATGTCGGCTGAGG 60  
DB 9 CCTCTGATGACCGGAGGAAGTTCTTTACCAATTCAAAATGTCGGCTGAGG 68  
QY 61 TCGGCGTGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120  
DB 69 TCGGCGTGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 128  
QY 121 ACTGACTTGGTTATCTTCCCAATTAAG 148  
DB 129 ACTGACTTGGTTATCTTCCCAATTAAG 156



Db 9 CCTTTGATGACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGGCTGGCTAAGG 68  
 QY 61 TCGGCGTGAGACCTACCTGCTGCTACGTAGTGAAGAGCGGTGACAGTCTACATCCTTTTC 120  
 Db 69 TCGGCGTGAGACCTACCTGCTGCTACGTAGTGAAGAGCGGTGACAGTCTACATCCTTTTC 128  
 QY 121 ACTGACCTTTGCTTATCTTCCCAATTAAG 148  
 Db 129 ACTGACCTTTGCTTATCTTCCCAATTAAG 156

RESULT 9  
 AF529821 596 bp mRNA linear PRI 17-SEP-2002  
 LOCUS Homo sapiens clone Ramos 7 AID (AID) mRNA, partial cds.  
 AF529821  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 596)  
 Martin, A. and Scharff, M.D.  
 Somatic hypermutation of the AID transgene in B and non-B cells  
 Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12304-12308 (2002)  
 2 (bases 1 to 596)  
 Martin, A. and Scharff, M.D.  
 Direct Submission  
 Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave. Bronx, NY 10461, USA

FEATURES  
 source  
 1..596  
 /location="Qualifiers"  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="Ramos 7"  
 1..>596  
 /gene="AID"  
 1..>596  
 /note="Integrated into Burkitt's lymphoma cell line Ramos"  
 /product="AID"  
 /protein\_id="AA095408.1"  
 /db\_xref="GI:22297230"  
 /translation="MDSILNMRKFLYQFNVRWAKGRRETYLCYVVKRDSATPSFLDFGILRNKNGHVELFLRYISDMDPDGRCYRTWFTSPCYDCARHVAADPLRGNPMLSRLITFARLYPCEDRKAPBEGRLRHARGVOQAIIMTFKDYICMNTFVNHRTFKAMEGHNHNSVRLSRQLRRIILPLYEVDLDRDAFTTLGL"

ORIGIN  
 Query Match 100.0%; Score 148; DB 9; Length 596;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-40;  
 Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 CCTTTGATGACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGGCTGGCTAAGG 60  
 QY 9 CCTTTGATGACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGGCTGGCTAAGG 68  
 Db 61 TCGGCGTGAGACCTACCTGCTGCTACGTAGTGAAGAGCGGTGACAGTCTACATCCTTTTC 120  
 QY 69 TCGGCGTGAGACCTACCTGCTGCTACGTAGTGAAGAGCGGTGACAGTCTACATCCTTTTC 128  
 Db 121 ACTGACCTTTGCTTATCTTCCCAATTAAG 148  
 QY 129 ACTGACCTTTGCTTATCTTCCCAATTAAG 156

RESULT 10  
 AF529824 596 bp mRNA linear PRI 17-SEP-2002  
 LOCUS Homo sapiens clone Ramos 10 AID (AID) mRNA, partial cds.  
 AF529824  
 DEFINITION

ACCESSION AF529824  
 VERSION AF529824.1 GI:22297235  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 596)  
 Martin, A. and Scharff, M.D.  
 Somatic hypermutation of the AID transgene in B and non-B cells  
 Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12304-12308 (2002)  
 2 (bases 1 to 596)  
 Martin, A. and Scharff, M.D.  
 Direct Submission  
 Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave. Bronx, NY 10461, USA

FEATURES  
 source  
 1..596  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="Ramos 10"  
 1..>596  
 /gene="AID"  
 1..>596  
 /note="Integrated into Burkitt's lymphoma cell line Ramos"  
 /product="AID"  
 /protein\_id="AA095411.1"  
 /db\_xref="GI:22297236"  
 /translation="MDSILNMRKFLYQFNVRWAKGRRETYLCYVVKRDSATPSFLDFGILRNKNGHVELFLRYISDMDPDGRCYRTWFTSPCYDCARHVAADPLRGNPMLSRLITFARLYPCEDRKAPBEGRLRHARGVOQAIIMTFKDYICMNTFVNHRTFKAMEGHNHNSVRLSRQLRRIILPLYEVDLDRDAFTTLGL"

ORIGIN  
 Query Match 100.0%; Score 148; DB 9; Length 596;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-40;  
 Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 CCTTTGATGACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGGCTGGCTAAGG 60  
 QY 9 CCTTTGATGACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGGCTGGCTAAGG 68  
 Db 61 TCGGCGTGAGACCTACCTGCTGCTACGTAGTGAAGAGCGGTGACAGTCTACATCCTTTTC 120  
 QY 69 TCGGCGTGAGACCTACCTGCTGCTACGTAGTGAAGAGCGGTGACAGTCTACATCCTTTTC 128  
 Db 121 ACTGACCTTTGCTTATCTTCCCAATTAAG 148  
 QY 129 ACTGACCTTTGCTTATCTTCCCAATTAAG 156

RESULT 11  
 AF529826 596 bp mRNA linear PRI 17-SEP-2002  
 LOCUS Homo sapiens clone Ramos 12 AID (AID) mRNA, partial cds.  
 AF529826  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 596)  
 Martin, A. and Scharff, M.D.  
 Somatic hypermutation of the AID transgene in B and non-B cells  
 Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12304-12308 (2002)  
 2 (bases 1 to 596)  
 Martin, A. and Scharff, M.D.  
 Direct Submission  
 Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of



Medicine, 1300 Morris Park Ave. Channing 404, Bronx, NY 10461, USA

FEATURES  
Location/Qualifiers  
1..596  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="Ramos 12"  
1..>596  
/gene="AID"  
1..>596  
/gene="AID"  
/note="Integrated into Burkitt's lymphoma cell line Ramos"  
/codon\_start=1  
/product="AID"  
/protein\_id="AAM95414.1"

CDS

ORIGIN

Query Match 100.0%; Score 148; DB 9; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.7e-40;  
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTCTGATGACCGGAGAGTTCTTTACCAATTGAAAGTCCGCTGGCTAAGG 60  
|||  
Db 9 CCTCTGATGACCGGAGAGTTCTTTACCAATTGAAAGTCCGCTGGCTAAGG 68

QY 61 TCGGCGTAGACCTACCTGCTGCTACGATGAGAGCGCTGACAGTCTCATCTTTTC 120  
|||  
Db 69 TCGGCGTAGACCTACCTGCTGCTACGATGAGAGCGCGTACAGTCTCATCTTTTC 128

QY 121 ACTGACCTTGGTTATCTTCGCATAAG 148  
|||  
Db 129 ACTGACCTTGGTTATCTTCGCATAAG 156

RESULT 12  
AF529827 596 bp mRNA linear PRI 17-SEP-2002  
LOCUS Homo sapiens clone Ramos 13 AID (AID) mRNA, partial cds.  
DEFINITION AF529827  
ACCESSION AF529827  
VERSION AF529827.1 GI:22297241  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 596)  
Martin, A. and Scharff, M.D.  
TITLE Somatic hypermutation of the AID transgene in B and non-B cells  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12304-12308 (2002)  
AUTHORS 2 (bases 1 to 596)  
Martin, A. and Scharff, M.D.  
TITLE Direct Submision  
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of  
Medicine, 1300 Morris Park Ave. Channing 404, Bronx, NY 10461, USA

FEATURES  
Location/Qualifiers  
1..596  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="Ramos 13"  
1..>596  
/gene="AID"  
1..>596  
/gene="AID"  
/note="Integrated into Burkitt's lymphoma cell line Ramos"  
/codon\_start=1  
/product="AID"  
/protein\_id="AAM95414.1"

/db\_xref="GI:22297242"  
/translation="MDSLMMRRKFLYQFNVMWAKRRRETYLCYVVKRDSATPSL  
DFGILRNKGCHEVLEFLRYISDMDDPGCRVMTFTWSPCYDCARHVDPLRGNP  
NLSLRIFARLYFCEDRKAPBEGRLRHRAQVQIAIMTFKDYFCWNTFVENHERTFK  
AMEGLHNSVRLSRQLRILLPLYEVDDLDAFRTLGL"

ORIGIN

Query Match 100.0%; Score 148; DB 9; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.7e-40;  
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTCTGATGACCGGAGAGTTCTTTACCAATTGAAAGTCCGCTGGCTAAGG 60  
|||  
Db 9 CCTCTGATGACCGGAGAGTTCTTTACCAATTGAAAGTCCGCTGGCTAAGG 68

QY 61 TCGGCGTAGACCTACCTGCTGCTACGATGAGAGCGCTGACAGTCTCATCTTTTC 120  
|||  
Db 69 TCGGCGTAGACCTACCTGCTGCTACGATGAGAGCGCGTACAGTCTCATCTTTTC 128

QY 121 ACTGACCTTGGTTATCTTCGCATAAG 148  
|||  
Db 129 ACTGACCTTGGTTATCTTCGCATAAG 156

RESULT 13  
AF529830 596 bp mRNA linear SYN 17-SEP-2002  
LOCUS Mus musculus clone 3 transgenic Homo sapiens AID (AID) mRNA,  
complete cds.  
DEFINITION AF529830  
ACCESSION AF529830  
VERSION AF529830.1 GI:22297247  
KEYWORDS  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 596)  
Martin, A. and Scharff, M.D.  
TITLE Somatic hypermutation of the AID transgene in B and non-B cells  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12304-12308 (2002)  
AUTHORS 2 (bases 1 to 596)  
Martin, A. and Scharff, M.D.  
TITLE Direct Submision  
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of  
Medicine, 1300 Morris Park Ave. Channing 404, Bronx, NY 10461, USA

FEATURES  
Location/Qualifiers  
1..596  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="3"  
/cell\_type="hybridoma PL-5"  
/transgenic  
1..596  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
1..>596  
/gene="AID"  
1..>596  
/gene="AID"  
/note="Integrated into Burkitt's lymphoma cell line Ramos"  
/codon\_start=1  
/product="AID"  
/protein\_id="AAM95417.1"  
/translation="MDSLMMRRKFLYQFNVMWAKRRRETYLCYVVKRDSATPSL  
DFGILRNKGCHEVLEFLRYISDMDDPGCRVMTFTWSPCYDCARHVDPLRGNP  
NLSLRIFARLYFCEDRKAPBEGRLRHRAQVQIAIMTFKDYFCWNTFVENHERTFK  
AMEGLHNSVRLSRQLRILLPLYEVDDLDAFRTLGL"

ORIGIN

Query Match 100.0%; Score 148; DB 12; Length 596;  
Best Local Similarity 100.0%; Pred. No. 1.7e-40;



	Matches	148;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	1	CCTCTTGATGACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGGCTGGGCTTAAGG	60							
Db	9	CCTTTGATGACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGGCTGGGCTTAAGG	68							
QY	61	TGGGCGTAGACCTACCTGTGCTACGTAGTAAGAGGCGTGACAGTGTACATCCTTTTC	120							
Db	69	TGGGCGTAGACCTACCTGTGCTACGTAGTAAGAGGCGTGACAGTGTACATCCTTTTC	128							
QY	121	ACTGACCTTTGGTTATCTTCGCAATAAG	148							
Db	129	ACTGACCTTTGGTTATCTTCGCAATAAG	156							

```

RESULT 14
AF529831
LOCUS
DEFINITION
AF529831 596 bp mRNA linear SYN 17-SEP-2002
Mus musculus clone 4 transgenic Homo sapiens AID (AID) mRNA,
complete cds.
AF529831
ACCESSION
AF529831.1 GI:22297249
VERSION
KEYWORDS
SOURCE
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 596)
AUTHORS Martin,A. and Schartf,M.D.
TITLE Somatic hypermutation of the AID transgene in B and non-B cells
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12304-12308 (2002)
AUTHORS
TITLE
JOURNAL
FEATURES
Location/Qualifiers
1..596

```

```

source
1..596
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1..596
/gene="AID"
1..596
/gene="AID"
/codon_start=1
/product="AID"
/protein_id="AAM95418.1"
/db_xref="GI:22297250"
/translation="MDSLNNRRKFLYQFKNVRWAKGRRETYLGVYKRRDASATSFSL
DGLRPNKNGCHVELLPLRYLISDMDIDGRGCRVYVWFSWSPCYDCAHVADPLRGNN
NLSRLIFARLYPCEDRKAKEPGLRRLHRAVQIALNTFKYIFYQNTTFVNHERTFP
AMEGHHENSVRLSRQFRILLPLVEVDLRLDAFRTGL"

```

	Query Match	100.0%	Score 148	DB 12	Length 596
	Best Local Similarity	100.0%	Pred. No. 1.7e-40		
	Matches 148	Conservative 0	Mismatches 0	Indels 0	Gaps 0
Oy	1 CCTCTGATGAACCGAGAAATTCTTACCAATTCAAATAATGTCGGGTGAAGG	60			
Db	9 CCTCTGATGAACCGAGAAATTCTTACCAATTCAAATAATGTCGGGTGAAGG	68			
Oy	61 TCGGCGTAGAAGCTACTGCTGCTACGTAGTGAAGAAGCGTGACAATGCTACATCCTTTTC	120			
Db	69 TCGGCGTAGAAGCTACTGCTGCTACGTAGTGAAGAAGCGTGACAATGCTACATCCTTTTC	128			

```

QY      121 ACTGACTTTGGTTATCTTGGCAATAAG 148
          |||||
Db      129 ACTGACTTTGGTTATCTTGGCAATAAG 156

```

RESULT 15  
AF529837  
LOCUS  
DEFINITION Mus musculus clone 10 transgenic homo sapiens AID (AID) mRNA,  
AF529837 complete cds.  
AF529837.1 GI:22297261  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
MUS musculus (house mouse)  
MUS musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 596)  
Martin,A. and Scharff,M.D.  
Somatic hypermutation of the AID transgene in B and non-B cells  
Proc. Natl. Acad. Sci. U.S.A. 99 (15), 12304-12308 (2002)  
2 (bases 1 to 596)  
Martin,A. and Scharff,M.D.  
Direct Submision  
Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of  
Medicine, 1300 Morris Park Ave. Bronx, NY 10461, USA  
Location/Qualifiers

```

source
1. .596
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/cclone="10"
/cell_type="hybridoma Pl-5"
/transgenic

source
1. .596
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1. .596
/gene="AID"
1. .596
/gene="AID"
/codon_start=1
/product="AID"
/protein_id="AA05424.1"
/translation="MDSILNRRKFFLYQFNKRWAKGRRETYLCYVKKRSDAITSFSLDGFGLRKGCHVEILPLRYISMDLDRGCRVTWTSMSPCYDCARHADFDGNPNPMSLRITATALYFCEDEKAPPEGRLRRHAGVCAIWTPKDYFCWMTFVNNHRTIKAMGSEHNSVRLSRQRLRLPLFVVDLDAFTTGL"

```

```

Query March 100.0% Score 148; DB 12; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.7e-40;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTCTTGATGACCGAGGAAAGTTCTTTACCAATTCAAAATATGTCGCGTGGCTTAAGG 60
Db 9 CCTCTTGATGACCGAGGAAAGTTCTTTACCAATTCAAAATATGTCGCGTGGCTTAAGG 68

QY 61 TCGGCGTGAGACCTACCTGTCGTACGTACGTAAAGGCGTGACAGTGTACATCCTTTTC 120
Db 69 TCGGCGTGAGACCTACCTGTCGTACGTACGTAAAGGCGTGACAGTGTACATCCTTTTC 128

QY 121 ACTGAGCTTGGTTATCTTCGCAATAAG 148
Db 129 ACTGAGCTTGGTTATCTTCGCAATAAG 156

Search completed: March 12, 2004, 21:22:30
Job time : 466.944 secs

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 12, 2004, 12:26:39 ; Search time 59.2641 Seconds  
(without alignments)  
10609.014 Million cell updates/sec

Title: US-09-966-880A-12

Perfect score: 148  
Sequence: 1 ccccttgatgaacagcagga.....ttggtatcttcgaataag 148

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_29Jan04:\*

- 1: Geneseq1980s:\*
- 2: Geneseq1990s:\*
- 3: Geneseq2000s:\*
- 4: Geneseq2001s:\*
- 5: Geneseq2001bs:\*
- 6: Geneseq2002s:\*
- 7: Geneseq2003as:\*
- 8: Geneseq2003bs:\*
- 9: Geneseq2003cs:\*
- 10: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	148	100.0	148	3 AAC55316	AAC55316 Human act
2	148	100.0	597	8 ADB88952	ADB88952 AID gene
3	148	100.0	1543	7 ABX05468	ABX05468 Human nov
4	148	100.0	2791	6 ABS73287	ABS73287 DNA encod
5	148	100.0	2791	6 ABS73288	ABS73288 DNA encod
6	148	100.0	2819	6 AAC55312	AAC55312 Human act
7	148	100.0	6564	3 AAC55314	AAC55314 Human act
8	148	100.0	11204	3 AAC55339	AAC55339 Human act
9	148	100.0	11204	6 ABS73286	ABS73286 DNA encod
10	109.6	74.1	2440	3 AAC55307	AAC55307 Mouse act
11	41.2	27.8	254	6 ABL80598	ABL80598 Human ova
12	41.2	27.8	1446	3 AAA12409	AAA12409 CDNA encd
13	41.2	27.8	1534	2 AAZ20856	AAZ20856 Polynucle
14	41.2	27.8	1534	4 AAS59293	AAS59293 Human CDN
15	41.2	27.8	1534	6 ABA90962	ABA90962 Human pol
16	41.2	27.8	1717	9 ADD18987	ADD18987 Human dis
17	41.2	27.8	20143	6 ABR35568	ABR35568 Human DNA
18	40.8	27.6	154	4 AAI27640	AAI27640 Probe #17
19	40.8	27.6	154	4 ABA75951	ABA75951 Human foe
20	40.8	27.6	154	4 AAI56609	AAI56609 Probe #25
21	40.8	27.6	154	4 ABA40515	ABA40515 Probe #18
22	40.8	27.6	154	4 AAK50628	AAK50628 Human bon
23	40.8	27.6	154	4 AAK24631	AAK24631 Human bra

24	40.8	27.6	154	4 ABR50231	ABR50231 Human liv
25	40.8	27.6	154	6 ABR24102	ABR24102 Human gen
26	40.8	27.6	300	2 AAZ13088	AAZ13088 Human gen
27	40.8	27.6	997	4 AAI18617	AAI18617 Probe #85
28	40.8	27.6	997	4 ABA63615	ABA63615 Human foe
29	40.8	27.6	997	4 AAI43731	AAI43731 Probe #12
30	40.8	27.6	997	4 ABA30812	ABA30812 Probe #92
31	40.8	27.6	997	4 AAK37857	AAK37857 Human bon
32	40.8	27.6	997	4 AAK12141	AAK12141 Human bra
33	40.8	27.6	997	4 ABR37486	ABR37486 Human liv
34	40.8	27.6	997	6 ABR11852	ABR11852 Human gen
35	40.8	27.6	1526	9 ADD18895	ADD18895 Human dis
36	40.8	27.6	1638	6 ABO61187	ABO61187 Unidentif
37	40.4	27.3	716	2 AAZ15848	AAZ15848 Human gen
38	39.4	26.6	371	8 ACH43025	ACH43025 Human foe
39	35.4	23.9	1348	6 ABR96785	ABR96785 Gene #328
40	32.4	21.9	2151	6 ADE62572	ADE62572 CDNA sequ
41	32.2	21.8	823	9 ADE25658	ADE25658 Human CDN
42	31.8	21.5	12222	4 AAF54045	AAF54045 Human alp
43	30	20.3	30	3 AAC55322	AAC55322 Human act
44	28.8	19.5	2529	4 ABL13752	ABL13752 Drosophil
45	28.8	19.5	2747	4 ABL13132	ABL13132 Drosophil

## ALIGNMENTS

RESULT 1	
AAC55316	
ID AAC55316 standard; DNA; 148 BP.	
AC AAC55316;	
XX	
DT 05-FEB-2001 (first entry)	
DE Human activation-induced cytidine deaminase exon 2 SEQ ID NO:12.	
XX	
KW Activation-induced cytidine deaminase; AID; cytidine deaminase;	
KW immune related disease; allergy; allergic disease; anti-allergic;	
KW antianemic; antiastrumatic; ophthalmological; anti-HIV; dermatological;	
KW gene therapy; B cell associated immune system disorder; food allergy;	
KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;	
KW drug nephritis; gamma-globulinemia; atopic dermatitis; allergic colitis;	
KW dry allergy; allergic rhinitis; Rosen disease; DiGeorge disease; AIDS;	
KW ataxia telangiectasia; common variable immunodeficiency disorder;	
KW major histocompatibility class II deficiency disease;	
KW auto immunodeficiency syndrome; IgG subclass selection disorder; ds.	
OS Homo sapiens.	
XX	
FN WO200058480-A1.	
XX	
PD 05-OCT-2000.	
XX	
PF 28-MAR-2000; 2000WO-JP001918.	
XX	
PR 29-MAR-1999; 99JP-00087192.	
PR 24-JUN-1999; 99JP-00178999.	
KW 27-DEC-1999; 99JP-00371382.	
XX	
PA (NISR) JAPAN TOBACCO INC.	
PA (HONJ) HONJO T.	
PI Honjo T, Muramatsu M;	
XX	
DR WPI; 2000-611715/58.	
XX	
PT Nucleic acid encoding activation induced cytidine deaminase, useful as a	
PT target for drug development for immune-related diseases including	
XX allergies.	
PS Claim 18; Page 150; 174pp; Japanese.	
XX	



CC The present invention describes an activation-induced cytidine deaminase  
CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and has  
CC cytidine activity similar to APOBEC-1. AID has anti-allergic, antineoplastic,  
CC antisthmatic, ophthalmological, anti-HIV and dermatological activities,  
CC and can be used in gene therapy. AID polynucleotides are useful in  
CC methods for identifying drugs for the treatment of B cell associated  
CC immune system disorders, immunodeficiency diseases and allergies, such as  
CC immunoglobulin A (IgA) deficiency disease, IgA nephritis, gamma-  
CC globulinemia, atopic dermatitis, allergic colitis, asthma, food allergy,  
CC drug allergy, allergic rhinitis, Rosen disease, Digorge disease, ataxia  
CC telangiectasia, common variable immunodeficiency disorder, MHC (major  
CC histocompatibility class II deficiency disease, AIDS (auto  
CC immunodeficiency syndrome), elevated IgE disorder, and IgG subclass  
CC selection disorder. The DNA sequences encoding AID may be used for gene  
CC therapy and the antibodies to the AID protein may be used for diagnosis  
CC and treatment of these disorders. The present sequence represents the  
CC exon 2 genomic DNA sequence of human AID

XX Sequence 148 BP; 35 A; 33 C; 38 G; 42 T; 0 U; 0 Other;

SQ Query Match 100.0%; Score 148; DB 3; Length 148;

Best Local Similarity 100.0%; Pred No. 1.7e-43; Mismatches 0; Indels 0; Gaps 0;

Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTCTGATGACCGAGGAGGAGTTCTTACCAATTCAAAATGTCGCTGAGCTAAGG 60

DB 1 CCTCTGATGACCGAGGAGGAGTTCTTACCAATTCAAAATGTCGCTGAGCTAAGG 60

QY 61 TCGGCGTAGACCTACCTGCTGCTACGTAAGAGGCGTGAAGTGTACATCCTTTTC 120

DB 61 TCGGCGTAGACCTACCTGCTGCTACGTAAGAGGCGTGAAGTGTACATCCTTTTC 120

QY 121 ACTGACCTTGTGTTATCTTGGCAATAAG 148

DB 121 ACTGACCTTGTGTTATCTTGGCAATAAG 148

#### RESULT 2

ADB88952 standard; DNA; 597 BP.

AC ADB88952;

DT 04-DEC-2003 (first entry)

DE AID gene as substrate used in AID mutagenic method.

XX ds; cytostatic; virucide; antidote; mutation;

KM activation-induced cytidine deaminase; cell phenotype;

KM monoclonal antibody; antigen; cross-reactivity; tumour;

KM germ warfare agent; toxin; V region.

XX Homo sapiens.

OS Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

XX Homo sapiens.

FT mutation

FT mutation

FT mutation

FT mutation

FT mutation

FT mutation

FT mutation

FT mutation

FT mutation

FT mutation

FT mutation

FT mutation

FT mutation

FT mutation

FT mutation

FT mutation

FT mutation

FT mutation

FT mutation

FT mutation

FT mutation

FT mutation

FT mutation

FT mutation

FT mutation

FT mutation

FT mutation

FT mutation

FT mutation

FT mutation

FT mutation

FT mutation

FT mutation

FT mutation

FT mutation

FT mutation

FT mutation

/replace(79,G)

/tag= g

/note= "mutation in Ramos cell"

/replace(85,G)

/tag= h

/note= "mutation in PI-5 hybridoma cell"

/replace(112,C)

/tag= i

/note= "mutation in Ramos cell"

/replace(118,G)

/tag= j

/note= "mutation in PI-5 hybridoma cell"

/replace(156,T)

/tag= k

/note= "mutation in PI-5 hybridoma cell"

/replace(161,A)

/tag= l

/note= "mutation in PI-5 hybridoma cell"

/replace(165,T,G)

/tag= m

/note= "mutation in Ramos cell, T mutation in CHO cell"

/replace(178,T)

/tag= n

/note= "mutation in Ramos cell"

/replace(206,G)

/tag= o

/note= "mutation in CHO cell"

/replace(209,C)

/tag= p

/note= "mutation in Ramos cell"

/replace(218,T)

/tag= q

/note= "mutation in Ramos cell"

/replace(250,C)

/tag= r

/note= "mutation in Ramos cell"

/replace(261,G)

/tag= s

/note= "mutation in Ramos cell"

/replace(289,A)

/tag= t

/note= "mutation in Ramos cell"

/replace(303,T)

/tag= u

/note= "mutation in Ramos cell"

/replace(329,T)

/tag= v

/note= "mutation in Ramos cell"

/replace(350,G,T)

/tag= w

/note= "T mutation in Ramos cell, G mutation in CHO cell"

/replace(356,A)

/tag= x

/note= "A mutation in Ramos cell, A mutation in CHO cell"

/replace(357,G)

/tag= y

/note= "mutation in Ramos cell"

/replace(367,T, A)

/tag= z

/note= "mutation in PI-5 hybridoma cell"

/replace(375,A)

/tag= aa

/note= "mutation in CHO cell"

/replace(402,A)

/tag= ab

/note= "mutation in PI-5 hybridoma cell"

/replace(410,T)

/tag= ac

/note= "mutation in PI-5 hybridoma cell"

/replace(418,T)

/tag= ad

/note= "mutation in PI-5 hybridoma cell"

/replace(420,T)

/note= "mutation in PI-5 hybridoma cell"



```
FT      /*tag= ae
FT      /note= "mutation in PI-5 hybridoma cell"
FT      replace(421,A)
FT      /*tag= af
FT      /note= "mutation in CHO cell"
FT      replace(430,G)
FT      /*tag= ag
FT      /note= "mutation in Ramos cell"
FT      replace(440,A)
FT      /*tag= ah
FT      /note= "mutation in CHO cell"
FT      replace(449,T)
FT      /*tag= ai
FT      /note= "mutation in PI-5 hybridoma cell"
FT      replace(495,A)
FT      /*tag= aj
FT      /note= "mutation in CHO cell"
FT      replace(517,C)
FT      /*tag= ak
FT      /note= "mutation in Ramos cell"
FT      replace(526,T, A)
FT      /*tag= al
FT      /note= "I. mutation in PI-5 hybridoma cell, A mutation in
FT      Ramos cell"
FT      replace(533,A)
FT      /*tag= am
FT      /note= "mutation in PI-5 hybridoma cell"
FT      replace(538,A)
FT      /*tag= an
FT      /note= "mutation in CHO cell"
FT      replace(549,A)
FT      /*tag= ao
FT      /note= "mutation in CHO cell"
FT      replace(568,T)
FT      /*tag= ap
FT      /note= "mutation in Ramos cell"
FT      replace(570,T)
FT      /*tag= aq
FT      /note= "mutation in CHO cell"
FT      /note= "mutation in CHO cell"
XX      WO2003061363-A2.
XX      31-JUL-2003.
XX      PD
XX      15-JAN-2003; 2003WO-US001149.
XX      PR
XX      17-JAN-2002; 2002US-0350269P.
XX      PA
XX      (YESH ) UNIV YESHIVA EINSTEIN COLLEGE.
XX      PI
XX      Martin A, Scharif MD;
XX      DR
XX      WPI; 2003-747983/70.
XX      PT
XX      Induction of mutations in gene expressed in eukaryotic cell useful e.g.
XX      to produce antibodies with greater affinity or specificity for antigens
XX      e.g. human pathogens by expressing an activation-induced cytidine
XX      deaminase gene in cell.
XX      PS
XX      Example 4; Fig 6; 73pp; English.
XX      CC
XX      The invention relates to the induction of mutations in a gene expressed
XX      in a eukaryotic cell, where the gene is operably linked to a promoter and
XX      within 2 kb of promoter, by expressing a transgenic activation-induced
XX      cytidine deaminase (AID) gene in the cell. The method is useful for
XX      producing mutated genes and proteins, determining the effect of a
XX      mutation on a protein or cell phenotype, and producing genes and proteins
XX      with altered properties. It is especially useful for producing antibodies
XX      with altered affinities or specificities for an antigen, or
XX      increased/decreased cross-reactivity for a second antigen, e.g. to
XX      produce antibodies with greater affinity/specificity useful
XX      therapeutically to treat tumours, viruses such as Ebola and lhasa Fever
XX      or against germ warfare agents, toxins (e.g. ricin). The method may
```

```
CC      similarly be used to alter affinity, specificity or cross-reactivity of a
CC      monoclonal antibody. This sequence represents the AID gene used in the
CC      method of the invention to induce mutations in the sequence.
XX      SQ
XX      Sequence 597 BP; 128 A; 164 C; 155 G; 150 T; 0 U; 0 Other;
XX      Query Match
XX      Best Local Similarity 100.0%; Score 148; DB 8; Length 597;
XX      Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX      1 CCTCTGATGAAACCGGAGAAAGTTCTTACCAATTCAAATATGTCCTGGGCTAAAGG 60
XX      9 CCTCTGATGAAACCGGAGAAAGTTCTTACCAATTCAAATATGTCCTGGGCTAAAGG 68
XX      61 TCGGCGTGAGACCTACCTGTGCTACGTAGTAAGAGCGGTGACAGTGTACATCCTTTTC 120
XX      69 TCGGCGTGAGACCTACCTGTGCTACGTAGTAAGAGCGGTGACAGTGTACATCCTTTTC 128
XX      QY
XX      121 ACTGACCTTTGGTTATCTTCGCAATAG 148
XX      129 ACTGACCTTTGGTTATCTTCGCAATAG 156
XX      Db
XX      RESULT 3
XX      ABRX05468
XX      ID ABRX05468 standard; cDNA; 1543 BP.
XX      AC ABRX05468;
XX      DT
XX      17-JAN-2003 (first entry)
XX      DE
XX      Human novel polynucleotide #483.
XX      KW
XX      Human; gene; ss; genetic disorder; gene mapping; medical imaging; cancer;
XX      neurodegenerative disorder; lymphoid cell disorder; osteoporosis;
XX      Parkinson's disease; Alzheimer's disease; bone degenerative disorder;
XX      osteoarthritis; periodontal disease; liver fibrosis; viral infection;
XX      fungal infection; bacterial infection; autoimmune disease; diabetes;
XX      atopic dermatitis.
XX      OS
XX      Homo sapiens.
XX      PN
XX      WO200274961-A1.
XX      PD
XX      26-SEP-2002.
XX      PF
XX      14-MAR-2002; 2002WO-US005109.
XX      PR
XX      15-MAR-2001; 2001US-00810173.
XX      PA
XX      (HYSE-) HYSEQ INC.
XX      PI
XX      Tang YT, Zhou P, Goodrich R, Asundi V, Zhang J, Zhao QH, Ren F;
XX      PI
XX      Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;
XX      PI
XX      Wehman T, Wang J, Wang D, Drmanac RT;
XX      DR
XX      WPI; 2003-040556/03.
XX      P-PSDB; ABRU00390.
XX      PT
XX      New isolated polypeptides and polynucleotides, useful for preventing,
XX      treating or ameliorating medical conditions, such as cancer,
XX      neurodegenerative disorders, lymphoid cell disorders, bone degenerative
XX      disorders, and infections.
XX      PS
XX      Claim 1; SEQ ID NO 483; 235pp; English.
XX      CC
XX      The invention relates to human polynucleotides and the polypeptides they
XX      encode. The polynucleotides and polypeptides are useful in diagnostics,
XX      forensics, gene mapping, medical imaging, identification of mutations
XX      responsible for genetic disorders or other traits, assessing biodiversity
XX      and producing many other types of data and products dependent on DNA and
XX      amino acid sequences. They are also useful for preventing, treating or
XX      ameliorating medical conditions, such as cancer, neurodegenerative
```



CC disorders (e.g. Parkinson's disease, Alzheimer's disease), lymphoid cell  
CC disorders, osteoporosis, osteoarthritis, bone degenerative disorders,  
CC peridontal disease, liver fibrosis, infections (e.g. viral, fungal or  
CC bacterial) or autoimmune diseases (e.g. diabetes, atopic dermatitis).  
CC Sequences ABX04986-ABX05511 represent human polynucleotides of the  
CC invention. Note: The sequence data for this patent is not represented in  
CC the printed specification but is based on sequence information supplied  
CC by the European Patent Office

XX  
SQ Sequence 1543 BP; 428 A; 334 C; 352 G; 429 T; 0 U; 0 Other;

Query Match 100.0%; Score 148; DB 7; Length 1543;  
Best Local Similarity 100.0%; Pred. No. 4e-43;

Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTCTGATGAGACCGGAGGAAGTTCTTACCAATTCAAAATGTCCTGGGCTAAAGG 60

DB 86 CCTCTGATGAGACCGGAGGAAGTTCTTACCAATTCAAAATGTCCTGGGCTAAAGG 145

QY 61 TCGGCGTGAAGACCTACCTGCTACGTAAGAGGCGTGAAGTCTACATCTTTTC 120

DB 146 TCGGCGTGAAGACCTACCTGCTACGTAAGAGGCGTGAAGTCTACATCTTTTC 205

QY 121 ACTGACTTGTGTTATCTTGCATTAAG 148

DB 206 ACTGACTTGTGTTATCTTGCATTAAG 233

RESULT 4  
AB573287  
ID AB573287 standard; DNA; 2791 BP.

AC AB573287;  
DT 04-DEC-2002 (first entry)

XX  
DE DNA encoding human translocation del(12p) protein #2.

XX Chromosome aberration; oncogenic fusion protein; cancer;  
XX proliferative disease; cellular protein isoform; heat shock protein 90;  
XX HSP-90; rheumatoid arthritis; cancer; haematopoietic disorder;  
XX T cell lymphoma; B cell lymphoma; chronic myeloid leukaemia; CML;  
XX acute myeloid leukaemia; AML; chronic myelomonocytic leukaemia; CMML;  
XX acute lymphoblastic leukaemia; ALL; APL; NHL; solid tumour;  
XX papillary thyroid carcinoma; Ewing's sarcoma; melanoma; liposarcoma;  
XX rhabdomyosarcoma; synovial sarcoma; viral infection; gene; ds.

XX Homo sapiens.

XX WO200269900-A2.

XX 12-SEP-2002.

XX 01-MAR-2002; 2002WO-US006518.

XX 01-MAR-2001; 2001US-0272751P.

XX (CONF-) CONFORMA THERAPEUTICS CORP.

XX Fritz LC, Burrows FJ;

XX WPI; 2002-698710/75.

XX P-PDB; ABG95083.

XX  
XX Treating genetically-defined disease associated with chromosomal  
XX aberrations yielding oncogenic fusion proteins, e.g. cell proliferative  
XX diseases, involves administering an inhibitor of heat shock protein 90.

XX  
XX Disclosure; Page 246-247; 389pp; English.

XX  
XX The invention describes a method of treating genetically-defined disease  
XX associated with chromosomal aberrations yielding oncogenic fusion  
XX proteins (I), treating cancerous cells containing (I) in a heterogeneous

CC cell population, treating proliferative diseases associated with mutant  
CC protein or cellular protein isoforms (II) dependent on heat shock protein  
CC (HSP)-90, or selectively treating cells expressing (II) involving  
CC administering HSP90-inhibitor. The method is useful for treating  
CC genetically-defined disease with chromosomal aberration yielding  
CC oncogenic fusion protein, treating cancerous cells containing fusion  
CC protein in heterogeneous cell population, treating proliferative disease  
CC (e.g. Rheumatoid arthritis or cancer) associated with mutant protein or  
CC cellular protein isoform dependent on heat shock protein (HSP)-90 (e.g.  
CC p53), or selectively treating cells expressing mutant protein or cellular  
CC protein isoform in a patient heterozygous for (II). The method is useful  
CC for treating a disease e.g. haematopoietic disorder such as T or B cell  
CC lymphoma, chronic myeloid leukaemia (CML), APL, ALL, NHL and CMML,  
CC or a disease characterised by a solid tumour such as papillary thyroid  
CC carcinoma, Ewing's sarcoma, melanoma, liposarcoma, rhabdomyosarcoma and  
CC synovial sarcoma. The method is also useful for treating viral  
CC infections. This represents the DNA sequence of a chromosome aberration

XX  
SQ Sequence 2791 BP; 842 A; 548 C; 625 G; 776 T; 0 U; 0 Other;

Query Match 100.0%; Score 148; DB 6; Length 2791;

Best Local Similarity 100.0%; Pred. No. 4.9e-43; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTCTGATGAGACCGGAGGAAGTTCTTACCAATTCAAAATGTCCTGGGCTAAAGG 60

DB 85 CCTCTGATGAGACCGGAGGAAGTTCTTACCAATTCAAAATGTCCTGGGCTAAAGG 144

QY 61 TCGGCGTGAAGACCTACCTGCTACGTAAGAGGCGTGAAGTCTACATCTTTTC 120

DB 145 TCGGCGTGAAGACCTACCTGCTACGTAAGAGGCGTGAAGTCTACATCTTTTC 204

QY 121 ACTGACTTGTGTTATCTTGCATTAAG 148

DB 205 ACTGACTTGTGTTATCTTGCATTAAG 232

RESULT 5  
AB573288  
ID AB573288 standard; DNA; 2791 BP.

AC AB573288;  
DT 04-DEC-2002 (first entry)

XX  
DE DNA encoding human translocation del(12p) protein #3.

XX Chromosome aberration; oncogenic fusion protein; cancer;  
XX proliferative disease; cellular protein isoform; heat shock protein 90;  
XX HSP-90; rheumatoid arthritis; cancer; haematopoietic disorder;  
XX T cell lymphoma; B cell lymphoma; chronic myeloid leukaemia; CML;  
XX acute myeloid leukaemia; AML; chronic myelomonocytic leukaemia; CMML;  
XX acute lymphoblastic leukaemia; ALL; APL; NHL; solid tumour;  
XX papillary thyroid carcinoma; Ewing's sarcoma; melanoma; liposarcoma;  
XX rhabdomyosarcoma; synovial sarcoma; viral infection; gene; ds.

XX Homo sapiens.

XX WO200269900-A2.

XX 12-SEP-2002.

XX 01-MAR-2002; 2002WO-US006518.

XX 01-MAR-2001; 2001US-0272751P.

XX (CONF-) CONFORMA THERAPEUTICS CORP.

XX Fritz LC, Burrows FJ;

XX WPI; 2002-698710/75.

XX P-PDB; ABG95084.



PT Treating genetically-defined disease associated with chromosomal  
 PT aberrations yielding oncogenic fusion proteins, e.g. cell proliferative  
 PT diseases, involves administering an inhibitor of heat shock protein 90.  
 XX  
 PS Disclosure, Page 248-249; 389pp; English.  
 XX  
 CC The invention describes a method of treating genetically-defined disease  
 CC associated with chromosomal aberrations yielding oncogenic fusion  
 CC proteins (I), treating cancerous cells containing (I) in a heterogeneous  
 CC cell population, treating proliferative diseases associated with mutant  
 CC protein or cellular protein isoforms (II) dependent on heat shock protein  
 CC (HSP)-90, or selectively treating cells expressing (II) involving  
 CC administering HSP90-inhibitor. The method is useful for treating  
 CC genetically-defined disease with chromosomal aberration yielding  
 CC oncogenic fusion protein, treating cancerous cells containing fusion  
 CC protein in heterogeneous cell population, treating proliferative disease  
 CC (e.g. rheumatoid arthritis or cancer) associated with mutant protein or  
 CC cellular protein isoform dependent on heat shock protein (HSP)-90 (e.g.  
 CC p53), or selectively treating cells expressing mutant protein or cellular  
 CC protein isoform in a patient heterozygous for (II). The method is useful  
 CC for treating a disease e.g. hematopoietic disorder such as T or B cell  
 CC lymphoma, chronic myeloid leukaemia (CML), APL, ALL, AML, NHL and CML,  
 CC or a disease characterized by a solid tumour such as papillary thyroid  
 CC carcinoma, Ewing's sarcoma, melanoma, liposarcoma, rhabdomyosarcoma and  
 CC synovial sarcoma. The method is also useful for treating viral  
 CC infections. This represents the DNA sequence of a chromosome aberration  
 CC  
 SQ Sequence 2791 BP; 842 A; 548 C; 625 G; 776 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 148; DB 6; Length 2791;  
 Best Local Similarity 100.0%; Pred. No. 4.9e-43;  
 Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CCTTTGATGACCGGAGAGATTCTTTACCAATTCAAAATGTCGCTGGGCTAAGG 60  
 DB CCTTTGATGACCGGAGAGATTCTTTACCAATTCAAAATGTCGCTGGGCTAAGG 144  
 QY 61 TCGGCGTAGACCTACCTGCTACGTAGTAGAAGGCGTAGACGTCTACATCCTTTTC 120  
 DB TCGGCGTAGACCTACCTGCTACGTAGTAGAAGGCGTAGACGTCTACATCCTTTTC 204  
 QY 121 ACTGACCTTGGTATCTTCGCAATTAAG 148  
 DB ACTGACCTTGGTATCTTCGCAATTAAG 232  
 RESULT 6  
 AACS5312  
 ID AACS5312 standard; cDNA; 2818 BP.  
 AC AACS5312;  
 XX  
 DT 05-FEB-2001 (first entry)  
 XX  
 DE Human activation-induced cytidine deaminase encoding cDNA SEQ ID NO:7.  
 XX  
 KW Activation-induced cytidine deaminase; AID; cytidine deaminase;  
 KW immune related disease; allergy; allergic disease; anti-allergic;  
 KW antineoplastic; antineoplastic; ophthalmological; anti-HIV; dermatological;  
 KW gene therapy; B cell associated immune system disorder; food allergy;  
 KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
 KW IGA nephritis; gamma-globulinemia; atopic dermatitis; allergic colitis;  
 KW drug allergy; allergic rhinitis; Rosen disease; DiGeorge disease; AIDS;  
 KW ataxia telangiectasia; common variable immunodeficiency disorder;  
 KW major histocompatibility class II deficiency disease;  
 KW auto immunodeficiency syndrome; Igg subclass selection disorder; ss.  
 XX  
 XX Homo sapiens.  
 OS  
 XX Key Location/Qualifiers  
 FH 80..676  
 FT CDS  
 FT /tag= a  
 FT /product= "activation-induced cytidine deaminase"

XX  
 PN WO200058480-A1.  
 XX  
 PD 05-OCT-2000.  
 XX  
 PF 28-MAR-2000; 2000WO-UP001918.  
 XX  
 PR 29-MAR-1999; 99JP-00087192.  
 PR 24-JUN-1999; 99JP-00178999.  
 PR 27-DEC-1999; 99JP-00371382.  
 XX  
 PA (NISH) JAPAN TOBACCO INC.  
 PA (HONB/) HONJO T.  
 PI Honjo T, Muramatsu M;  
 XX  
 DR MPI: 2000-611715/58.  
 DR P-PSDB; AAB24198.  
 XX  
 PT Nucleic acid encoding activation induced cytidine deaminase, useful as a  
 PT target for drug development for immune-related diseases including  
 PT allergies.  
 XX  
 PS Claim 3; Page 135-139; 174pp; Japanese.  
 XX  
 CC The present sequence encodes human activation-induced cytidine deaminase  
 CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and has  
 CC cytidine activity similar to APOBEC-1. AID has antiallergic, antineoplastic,  
 CC antineoplastic, ophthalmological, anti-HIV and dermatological activities,  
 CC and can be used in gene therapy. AID polynucleotides are useful in  
 CC methods for identifying drugs for the treatment of B cell associated  
 CC immune system disorders, immunodeficiency diseases and allergies, such as  
 CC immunoglobulin A (IgA) deficiency disease, IGA nephritis, gamma-  
 CC globulinemia, atopic dermatitis, allergic colitis, asthma, food allergy,  
 CC drug allergy, allergic rhinitis, Rosen disease, DiGeorge disease, ataxia  
 CC telangiectasia, common variable immunodeficiency disorder, MHC (major  
 CC histocompatibility class) class II deficiency disease, AIDS (auto  
 CC immunodeficiency syndrome), elevated Igg disorder, and Igg subclass  
 CC selection disorder. The DNA sequences encoding AID may be used for gene  
 CC therapy and the antibodies to the AID protein may be used for diagnosis  
 CC and treatment of these disorders  
 CC  
 SQ Sequence 2818 BP; 868 A; 548 C; 626 G; 776 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 148; DB 3; Length 2818;  
 Best Local Similarity 100.0%; Pred. No. 4.9e-43;  
 Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CCTTTGATGACCGGAGAGATTCTTTACCAATTCAAAATGTCGCTGGGCTAAGG 60  
 DB CCTTTGATGACCGGAGAGATTCTTTACCAATTCAAAATGTCGCTGGGCTAAGG 147  
 QY 61 TCGGCGTAGACCTACCTGCTACGTAGTAGAAGGCGTAGACGTCTACATCCTTTTC 120  
 DB TCGGCGTAGACCTACCTGCTACGTAGTAGAAGGCGTAGACGTCTACATCCTTTTC 207  
 QY 121 ACTGACCTTGGTATCTTCGCAATTAAG 148  
 DB ACTGACCTTGGTATCTTCGCAATTAAG 235  
 RESULT 7  
 AACS5314  
 ID AACS5314 standard; DNA; 6564 BP.  
 AC AACS5314;  
 XX  
 DT 05-FEB-2001 (first entry)  
 XX  
 DE Human activation-induced cytidine deaminase genomic DNA SEQ ID NO:10.  
 XX  
 KW Activation-induced cytidine deaminase; AID; cytidine deaminase;  
 KW immune related disease; allergy; allergic disease; anti-allergic;



KM antianaemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;  
 KM gene therapy; B cell associated immune system disorder; food allergy;  
 KM immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
 KM IGA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;  
 KM drug allergy; allergic rhinitis; Rosen disease; Digorge disease; AIDS;  
 KM ataxia telangiectasia; common variable immunodeficiency disorder;  
 KM major histocompatibility class II deficiency disease;  
 KM auto immunodeficiency syndrome; Igg subclass selection disorder; ds.  
 OS Homo sapiens.  
 XX  
 PN WO200058480-A1.  
 XX  
 PD 05-OCT-2000.  
 XX  
 PF 28-MAR-2000; 2000WO-JP001918.  
 XX  
 PR 29-MAR-1999; 99JP-00087192.  
 PR 24-JUN-1999; 99JP-00178999.  
 PR 27-DEC-1999; 99JP-00371382.  
 XX  
 PA (NISR) JAPAN TOBACCO INC.  
 PA (HONJ/) HONJO T.  
 XX  
 PI Honjo T, Muramatsu M;  
 XX  
 DR WPI; 2000-611715/58.  
 XX  
 XX Nucleic acid encoding activation induced cytidine deaminase, useful as a  
 PT target for drug development for immune-related diseases including  
 PT allergies.  
 XX  
 PS Claim 17; Page 145-150; 174pp; Japanese.  
 XX  
 CC The present invention describes an activation-induced cytidine deaminase  
 CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and has  
 CC cytidine activity similar to APOBEC-1. AID has antiasthmatic, antianaemic,  
 CC antiasthmatic, ophthalmological, anti-HIV and dermatological activities;  
 CC and can be used in gene therapy. AID polynucleotides are useful in  
 CC methods for identifying drugs for the treatment of B cell associated  
 CC immune system disorders, immunodeficiency diseases and allergies, such as  
 CC immunoglobulin A (IGA) deficiency disease, IGA nephritis, gamma-  
 CC globulinaemia, atopic dermatitis, allergic colitis, asthma, food allergy,  
 CC drug allergy, allergic rhinitis, Rosen disease, Digorge disease, ataxia  
 CC telangiectasia, common variable immunodeficiency disorder, MHC (major  
 CC histocompatibility class) class II deficiency disease, AIDS (auto  
 CC immunodeficiency syndrome), elevated Igg disorder, and Igg subclass  
 CC selection disorder. The DNA sequences encoding AID may be used for gene  
 CC therapy and the antibodies to the AID protein may be used for diagnosis  
 CC and treatment of these disorders. The present sequence represents a  
 CC genomic DNA sequence of human AID  
 CC  
 XX  
 SQ Sequence 6564 BP; 1909 A; 1358 C; 1383 G; 1914 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 148; DB 3; Length 6564;  
 Best Local Similarity 100.0%; Pred. No. 6.7e-43;  
 Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ID AAC55339 standard; DNA; 11204 BP.  
 XX  
 AC AAC55339;  
 XX  
 DT 05-FEB-2001 (first entry)  
 XX  
 DE Human activation-induced cytidine deaminase genomic DNA SEQ ID NO:35.  
 KM Activation-induced cytidine deaminase; AID; cytidine deaminase;  
 KM immune related disease; allergy; allergic disease; antiasthmatic;  
 KM antianaemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;  
 KM gene therapy; B cell associated immune system disorder; food allergy;  
 KM immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
 KM IGA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;  
 KM drug allergy; allergic rhinitis; Rosen disease; Digorge disease; AIDS;  
 KM ataxia telangiectasia; common variable immunodeficiency disorder;  
 KM major histocompatibility class II deficiency disease;  
 KM auto immunodeficiency syndrome; Igg subclass selection disorder; ds.  
 OS Homo sapiens.  
 XX  
 PN WO200058480-A1.  
 XX  
 PD 05-OCT-2000.  
 XX  
 PF 28-MAR-2000; 2000WO-JP001918.  
 XX  
 PR 29-MAR-1999; 99JP-00087192.  
 PR 24-JUN-1999; 99JP-00178999.  
 PR 27-DEC-1999; 99JP-00371382.  
 XX  
 PA (NISR) JAPAN TOBACCO INC.  
 PA (HONJ/) HONJO T.  
 XX  
 PI Honjo T, Muramatsu M;  
 XX  
 DR WPI; 2000-611715/58.  
 XX  
 XX Nucleic acid encoding activation induced cytidine deaminase, useful as a  
 PT target for drug development for immune-related diseases including  
 PT allergies.  
 XX  
 PS Claim 17; Page 163-170; 174pp; Japanese.  
 XX  
 CC The present invention describes an activation-induced cytidine deaminase  
 CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and has  
 CC cytidine activity similar to APOBEC-1. AID has antiasthmatic, antianaemic,  
 CC antiasthmatic, ophthalmological, anti-HIV and dermatological activities;  
 CC and can be used in gene therapy. AID polynucleotides are useful in  
 CC methods for identifying drugs for the treatment of B cell associated  
 CC immune system disorders, immunodeficiency diseases and allergies, such as  
 CC immunoglobulin A (IGA) deficiency disease, IGA nephritis, gamma-  
 CC globulinaemia, atopic dermatitis, allergic colitis, asthma, food allergy,  
 CC drug allergy, allergic rhinitis, Rosen disease, Digorge disease, ataxia  
 CC telangiectasia, common variable immunodeficiency disorder, MHC (major  
 CC histocompatibility class) class II deficiency disease, AIDS (auto  
 CC immunodeficiency syndrome), elevated Igg disorder, and Igg subclass  
 CC selection disorder. The DNA sequences encoding AID may be used for gene  
 CC therapy and the antibodies to the AID protein may be used for diagnosis  
 CC and treatment of these disorders. The present sequence represents a  
 CC genomic DNA sequence of human AID  
 CC  
 XX  
 SQ Sequence 11204 BP; 3305 A; 2273 C; 2373 G; 3253 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 148; DB 3; Length 11204;  
 Best Local Similarity 100.0%; Pred. No. 8.1e-43;  
 Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



Db 6340 TCGGCGTGAACCTTGTCTAGTGAAGAGCGTGACAGTCTCATCTTTTC 6399  
 QY 121 ACTGACCTTGGTATCTTGGCATTAAG 148  
 Db 6400 ACTGACCTTGGTATCTTGGCATTAAG 6427

RESULT 9  
 ABS73286  
 ID ABS73286 standard; DNA; 11204 BP.

XX ABS73286;  
 AC 04-DEC-2002 (first entry)  
 DT  
 XX DNA encoding human translocation del(12p) protein #1.

XX Chromosome aberration; oncogenic fusion protein; cancer;  
 KW proliferative disease; cellular fusion isoform; heat shock protein 90;  
 KW HSP-90; rheumatoid arthritis; cancer; haematopoietic disorder;  
 KW T cell lymphoma; B cell lymphoma; chronic myeloid leukaemia; CML;  
 KW acute myeloid leukaemia; AML; chronic myelomonocytic leukaemia; CMML;  
 KW acute lymphoblastic leukaemia; ALL; APL; NHL; solid tumour;  
 KW papillary thyroid carcinoma; Ewing's sarcoma; melanoma; liposarcoma;  
 KW rhabdomyosarcoma; synovial sarcoma; viral infection; gene; ds.

XX Homo sapiens.  
 OS  
 XX WO200269900-A2.

XX 12-SEP-2002.  
 XX 01-MAR-2002; 2002WO-US006518.  
 XX 01-MAR-2001; 2001US-0272751P.

XX (CONF-) CONFORMA THERAPEUTICS CORP.

XX Filtz LC, Burrows FU;  
 XX WPI: 2002-696710/75.  
 DR P-PSDB; ABG95082.

XX Treating genetically-defined disease associated with chromosomal  
 PT aberrations yielding oncogenic fusion proteins, e.g. cell proliferative  
 PT diseases, involves administering an inhibitor of heat shock protein 90.  
 XX  
 PS Disclosure; Page 242-245; 389pp; English.

XX The invention describes a method of treating genetically-defined disease  
 CC associated with chromosomal aberrations yielding oncogenic fusion  
 CC proteins (I), treating cancerous cells containing (I) in a heterogeneous  
 CC cell population, treating proliferative diseases associated with mutant  
 CC protein or cellular protein isoforms (II) dependent on heat shock protein  
 CC (HSP)-90, or selectively treating cells expressing (II) involving  
 CC administering HSP90-inhibitor. The method is useful for treating  
 CC genetically-defined disease with chromosomal aberration yielding  
 CC oncogenic fusion protein, treating cancerous cells containing fusion  
 CC protein in heterogeneous cell population, treating proliferative disease  
 CC (e.g. Rheumatoid arthritis or cancer) associated with mutant protein or  
 CC cellular protein isoform dependent on heat shock protein (HSP)-90 (e.g.  
 CC P53), or selectively treating cells expressing mutant protein or cellular  
 CC protein isoform in a patient heterozygous for (II). The method is useful  
 CC for treating a disease e.g. haematopoietic disorder such as T or B cell  
 CC lymphoma, chronic myeloid leukaemia (CML), APL, ALL, AML, NHL and CMML,  
 CC or a disease characterised by a solid tumour such as papillary thyroid  
 CC carcinoma, Ewing's sarcoma, melanoma, liposarcoma, rhabdomyosarcoma and  
 CC synovial sarcoma. The method is also useful for treating viral  
 CC infections. This represents the DNA sequence of a chromosome aberration  
 XX  
 SQ Sequence 11204 BP; 3305 A; 2273 C; 2373 G; 3253 T; 0 U; 0 Other;

Query Match 100.0%; Score 148; DB 6; Length 11204;  
 Best Local Similarity 100.0%; Pred. No. 8.1e-43;  
 Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTCTGATGAACCGGAGAAATTTCTTACCAATTCAAAATGCCGCTGGCTAAAGG 60  
 Db 6280 CCTCTGATGAACCGGAGAAATTTCTTACCAATTCAAAATGCCGCTGGCTAAAGG 6339  
 QY 61 TCGGCGTGAACCTTGTCTAGTGAAGAGCGTGACAGTCTCATCTTTTC 120  
 Db 6340 TCGGCGTGAACCTTGTCTAGTGAAGAGCGTGACAGTCTCATCTTTTC 6399

QY 121 ACTGACCTTGGTATCTTGGCATTAAG 148  
 Db 6400 ACTGACCTTGGTATCTTGGCATTAAG 6427

RESULT 10  
 AAC5307  
 ID AAC5307 standard; cDNA; 2440 BP.

XX AAC5307;  
 AC  
 DT 05-FEB-2001 (first entry)

DE Mouse activation-induced cytidine deaminase encoding cDNA SEQ ID NO:1.

XX Activation-induced cytidine deaminase; AID; cytidine deaminase;  
 KW immune related disease; allergy; allergic disease; antiallergic;  
 KW antianemic; antisthmatic; ophthalmological; anti-HIV; dermatological;  
 KW gene therapy; B cell associated immune system disorder; food allergy;  
 KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
 KW IGA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;  
 KW drug allergy; allergic rhinitis; Rosen disease; DiGeorge disease; AIDS;  
 KW ataxia telangiectasia; common variable immunodeficiency disorder;  
 KW major histocompatibility class II deficiency disease;  
 KW auto immunodeficiency syndrome; IgG subclass selection disorder; ss.

XX Mus musculus.

XX Key Location/Qualifiers  
 FT CDS 93..689  
 FT /\*tag= a  
 FT /product= "activation-induced cytidine deaminase"

XX MO200058480-A1.  
 XX  
 PD 05-OCT-2000.

XX 28-MAR-2000; 2000WO-TP001918.

XX 29-MAR-1999; 99JP-00087192.  
 PR 24-JUN-1999; 99JP-00178999.  
 PR 27-DEC-1999; 99JP-00371382.

XX (NISE) JAPAN TOBACCO INC.  
 PA (HONO/) HONO T.

PI Honjo T, Muramatsu M;  
 XX WPI: 2000-611715/58.  
 DR P-PSDB; AAB24197.

XX Nucleic acid encoding activation induced cytidine deaminase, useful as a  
 PT target for drug development for immune-related diseases including  
 PT allergies.

XX Claim 3; Page 126-130; 174pp; Japanese.

CC The present sequence encodes mouse activation-induced cytidine deaminase  
 CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and has  
 CC cytidine activity similar to APOBEC-1. AID has antiallergic, antianemic,  
 CC antisthmatic, ophthalmological, anti-HIV and dermatological activities.



CC and can be used in gene therapy. AID polynucleotides are useful in  
 CC methods for identifying drugs for the treatment of B cell associated  
 CC immune system disorders, immunodeficiency diseases and allergies, such as  
 CC immunoglobulin A (IGA) deficiency disease, IGA nephritis, gamma-  
 CC globulinemia, atopic dermatitis, allergic colitis, asthma, food allergy,  
 CC drug allergy, allergic rhinitis, rosacea disease, Disgeorge disease, ataxia  
 CC telangiectasia, common variable immunodeficiency disorder, MHC (major  
 CC histocompatibility class II deficiency disease, AIDS (auto  
 CC immunodeficiency syndrome), elevated IgE disorder, and IgG subclass  
 CC selection disorder. The DNA sequences encoding AID may be used for gene  
 CC therapy and the antibodies to the AID protein may be used for diagnosis  
 CC and treatment of these disorders  
 CC

XX Sequence 2440 BP; 706 A; 546 C; 551 G; 636 T; 0 U; 1 Other;

Query Match 74.1%; Score 109.6; DB 3; Length 2440;

Best Local Similarity 83.8%; Pred. No. 4e-29; Mismatches 24; Indels 0; Gaps 0;

Matches 124; Conservative 0; Mismatches 24; Indels 0; Gaps 0;  
 QY 1 CCTCTGATGAACCGAGAGAGTTCTTACCAATTCAAAATGTCGCGTGGGCTAAGG 60  
 DB 101 CCTTCTATGAGAGAAAGAGTTCTTACCAATTCAAAATGTCGCGTGGGCTAAGG 160  
 QY 61 TCGGCGTAGACCTACCTGCTGCTAGTGAAGAGCGCTGACAGTCTACATCTTTTC 120  
 DB 161 ACGGATGAGACCTACCTGCTGCTAGTGAAGAGAGATAGTCCACTCTGCTC 220  
 QY 121 ACTGAGCTTGTGTATCTTGGCAATAG 148  
 DB 221 ACTGAGCTTGGCCACTTGGCAACAG 248

#### RESULT 11

ABE80598  
 ID ABL80598 standard; cDNA; 254 BP.

XX ABL80598;

XX 17-MAY-2002 (first entry)

XX Human ovarian cancer related cDNA clone SEQ ID NO:3576.

XX Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.

XX Homo sapiens.

XX WO200192581-A2.

XX 06-DEC-2001.

XX 29-MAY-2001; 2001WO-US017756.

XX 26-MAY-2000; 2000US-0207484P.

XX (CORI-) CORIXA CORP.

XX Algate PA, Harlocker SL, Jones R;

XX WPI; 2002-122075/16.

XX Composition for therapy and diagnosis of ovarian cancer comprising  
 PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding  
 PT polypeptide, antibody specific to polypeptide or T cell expressing  
 PT polypeptide.

XX Claim 1; SEQ ID NO 3576; 469pp; English.

XX The present invention describes a composition (I) comprising: carriers  
 CC and immunostimulants; and a polypeptide (II) of a ovarian tumour  
 CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence (SI)  
 CC from the 10912 nucleotide sequences as given in ABL77023 to ABL87934,  
 CC (III) encoding (II) having a sequence (S2), a T cell population of (II),  
 CC or antigen presenting cells that express (II). (I) has cytostatic

CC activity. An oligonucleotide (IV) that hybridises to (SI) can be used for  
 CC detecting ovarian cancer in a patient's biological sample preferably  
 CC serum or ovarian tissue. The method comprises contacting a biological  
 CC sample from a patient with (IV), detecting the amount of polynucleotide  
 CC hybridising to (IV) and comparing the amount to a predetermined cutoff  
 CC value and thereby detecting ovarian cancer in the patient, where the  
 CC amount of polynucleotide hybridising to (IV) is detected preferably by  
 CC polymerase chain reaction (PCR). (I) comprising (III) and/or (II) is  
 CC useful for stimulating and/or expanding T cells specific for an ovarian  
 CC tumour protein comprising contacting T cells with (III) or (II). (III) is  
 CC useful in design and preparation of ribozyme molecules for inhibiting  
 CC expression of the tumour polypeptides and proteins in tumour cells; and  
 CC to isolate a full length gene from a suitable library e.g., a tumour cDNA  
 CC library using well known techniques  
 CC

XX Sequence 254 BP; 57 A; 66 C; 65 G; 59 T; 0 U; 7 Other;

Query Match 27.8%; Score 41.2; DB 6; Length 254;

Best Local Similarity 57.9%; Pred. No. 0.00012; Mismatches 53; Indels 0; Gaps 0;

QY 2 CTTGTGATGAACCGAGAGAGTTCTTACCAATTCAAAATGTCGCGTGGGCTAAGG 61  
 DB 7 CACTGATGATCCACCCACATTCATTCAACTTAACATGACCTTGGGTACAGGA 66  
 QY 62 CGGCGTAGACCTACCTGCTGCTAGTGAAGAGCGCTACAGTCTACATCTTTTCA 121  
 DB 67 CGGATGAGACTTACCTGCTGCTAGTGAAGAGCGCTACAGTCTACATCTTTTCA 126  
 QY 122 CTGAGC 127  
 DB 127 CTGAAC 132

#### RESULT 12

AAA12409  
 ID AAA12409 standard; cDNA; 1446 BP.

XX AAA12409;

XX 25-JUL-2000 (first entry)

XX cDNA encoding a human RNA-associated protein.

XX Human; RNA-associated protein; cell proliferation; cancer; inflammation;  
 KW immune response; reproductive disorder; actinic keratosis;  
 KW atherosclerosis; arteriosclerosis; bursitis; cirrhosis; hepatitis;  
 KW mixed connective tissue disease; myelofibrosis; primary thrombocythemia;  
 KW paroxysmal nocturnal hemoglobinuria; polycythemia vera; psoriasis;  
 KW trauma; ss.

XX Homo sapiens.

XX Key

XX CDS

XX Location/Qualifiers

XX 282..1446

XX /tag= a

XX /product= "RNA-associated protein"

XX WO200015799-A2.

XX 23-MAR-2000.

XX 17-SEP-1999; 99WO-US021688.

XX 17-SEP-1998; 98US-00156039.

XX 22-SEP-1998; 98US-00158720.

XX 04-NOV-1998; 98US-00186415.

XX 08-APR-1999; 99US-0128660P.

XX (INCY-) INCYTE PHARM INC.

XX Tang YT, Corley NC, Guegler KJ, Gorgone GA, Patterson C;

XX Hillman JL, Baughn MR, Lal P, Azimzal Y, Yue H, Yang J;



XX WPI: 2000-271437/23.  
 DR P-PDB; AAY84437.  
 XX  
 PT New polypeptides and polynucleotides, useful for preventing and treating  
 PT a disorder associated with increased or decreased expression of RNA  
 PT associated proteins.  
 XX  
 PS Claim 9; Page 119; 131pp; English.  
 XX  
 CC The present sequence encodes a human RNA-associated protein. The  
 CC expression of RNA-associated proteins is closely associated with  
 CC reproductive tissues, nervous tissues, cell proliferation including  
 CC cancer, inflammation and immune responses, and so they may be used for  
 CC diagnosis, treatment or prevention of cell proliferative, diseases and  
 CC immune/inflammatory disorders, and reproductive disorders. Diseases and  
 CC disorders which may be treated include actinic keratosis,  
 CC atherosclerosis, arteriosclerosis, bursitis, cirrhosis, hepatitis, mixed  
 CC connective tissue disease, myelofibrosis, paroxysmal nocturnal  
 CC hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia  
 CC and cancers, and trauma  
 XX  
 SQ Sequence 1446 BP; 370 A; 374 C; 393 G; 308 T; 0 U; 1 Other;  
 Query Match 27.8%; Score 41.2; DB 3; Length 1446;  
 Best Local Similarity 57.9%; Pred. No. 0.00022;  
 Matches 73; Conservative 0; Mismatches 53; Indels 0; Gaps 0;  
 QY 2 CTCTTGATGAACCGAGAGATTCTTTACCAATTCAAAATGTCGCTGGCTAAGGT 61  
 DB 864 CACTCGATGATCCACCACTTACTTCACTTTACATGACCTTGGTCAGAGA 923  
 QY 62 CGGCGTAGACCTTCTGTGTACGTAGTGAAGGCGGTGACAGTGTACATCTTTCA 121  
 DB 924 CGGATGAGACTTCTGTGTATGAGGTGAGGCGCATGACATGACCTGGTCTTG 983  
 QY 122 CTGAC 127  
 DB 984 CTGAC 989  
 RESULT 13  
 AA220856  
 ID AA220856 standard; cDNA; 1534 BP.  
 XX  
 AC AA220856;  
 XX  
 DT 09-DEC-1999 (first entry)  
 XX  
 DE Polynucleotide sequence of the 1p547\_4 clone.  
 XX  
 KW secreted protein; cDNA library; clone; transmembrane protein;  
 KM signal sequence cloning; hybridization cloning; gene therapy; receptor;  
 KM ds.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 51..1205  
 FT /\*tag= a  
 FT /product= "1p547\_4"  
 FT /\*tag= b  
 FT /\*tag= 981..1205  
 FT mat\_peptide  
 FT /\*tag= c  
 XX  
 PN W09942470-A1.  
 XX  
 PD 26-AUG-1999.  
 XX  
 PF 18-FEB-1999; 99WO-US003458.  
 PF 18-FEB-1998; 98US-0075038P.  
 PR

PR 17-FEB-1999; 99US-00251600.  
 XX  
 XX (GENY ) GENETICS INST INC.  
 PA  
 XX  
 PI Jacobs K, McCoy JM, Lavalie ER, Collins-Racie LA, Merberg D;  
 PI Treacy M, Agostino MJ, Steininger RJ;  
 XX  
 DR WPI: 1999-518580/43.  
 DR P-PDB; AAY42383.  
 XX  
 PT New polynucleotides encoding human secreted proteins used for  
 PT therapeutic, diagnostic and research purposes.  
 XX  
 PS Claim 16; Page 104; 125pp; English.  
 XX  
 CC This is the polynucleotide sequence of the clone 1p547\_4, which was  
 CC isolated from a human fetal brain cDNA library using methods which are  
 CC selective for cDNAs encoding secreted proteins, or by identification as a  
 CC secreted or transmembrane protein on the basis of computer analysis of  
 CC the amino acid sequence of the encoded protein. The pns and proteins of  
 CC the invention are predicted to have biological activities which would  
 CC make them suitable for treating, preventing or ameliorating medical  
 CC conditions in humans and animals, although no supporting data is given.  
 CC Suggested activities include nutritional activity, cytokine and cell  
 CC proliferation/differentiation activity, immune stimulating (e.g. as  
 CC vaccines) or suppressing activity, hematopoiesis regulating activity,  
 CC tissue growth activity, activin/inhibin activity, thrombolytic activity,  
 CC chemotactic/chemokinetic activity, hemostatic and thrombotic activity,  
 CC receptor/ligand activity, anti-inflammatory activity, cadherin/tumor  
 CC invasion suppressor activity, and tumor inhibition activity. The pns are  
 CC also stated to be useful for gene therapy. Other activities include  
 CC inhibiting the growth, infection or function of bacteria, fungi, viruses  
 CC and other parasites; effecting bodily characteristics such as, e.g.  
 CC weight, color, skin, etc., effecting biorythms or circadian cycles;  
 CC enhancing fertility; treatment of depression; treatment of pain; hormonal  
 CC or endocrine activity  
 XX  
 SQ Sequence 1534 BP; 495 A; 363 C; 345 G; 331 T; 0 U; 0 Other;  
 Query Match 27.8%; Score 41.2; DB 2; Length 1534;  
 Best Local Similarity 57.9%; Pred. No. 0.00022;  
 Matches 73; Conservative 0; Mismatches 53; Indels 0; Gaps 0;  
 QY 2 CTCTTGATGAACCGAGAGATTCTTTACCAATTCAAAATGTCGCTGGCTAAGGT 61  
 DB 633 CACTCGATGATCCACCACTTCACTTTACATGACCTTGGTCAAGAGA 692  
 QY 62 CGGCGTAGACCTTCTGTGTACGTAGTGAAGGCGGTGACAGTGTACATCTTTCA 121  
 DB 693 CGGATGAGACTTCTGTGTATGAGGTGAGGCGCATGACATGACCTGGTCTTG 752  
 QY 122 CTGAC 127  
 DB 753 CTGAC 758  
 RESULT 14  
 AAS59293  
 ID AAS59293 standard; cDNA; 1534 BP.  
 XX  
 AC AAS59293;  
 XX  
 DT 16-JAN-2002 (first entry)  
 XX  
 DE Human cDNA encoding a secreted protein 1p547\_4.  
 XX  
 KW Human; secreted protein; sr; antiinflammatory; immunosuppressive;  
 KW nootropic; neuroprotective; antiarthritic; antimicrobial; vlnetary;  
 KW cytostatic; antidiabetic; virocid; antiinfectivity; anticonvulsant;  
 KW vasotropic; antiparkinsonian; immunostimulant; dermatological;  
 KW antineumatic; antitumor; antitumor; osteopathic; tranquiliser;  
 KW cerebroprotective; cytokine; cell proliferation; cell differentiation;  
 KW immune deficiency; severe combined immunodeficiency; SCID; tumour;  
 KW



XX autoimmune disorder; multiple sclerosis; rheumatoid arthritis;  
KW graft-versus-host disease; myeloid deficiency; wound healing; ulcer;  
KW periodontal disease; osteoporosis; osteoarthritis; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease; infection; cardiac disease;  
KW stroke; sepsis; inflammatory bowel disease; contraceptive; immunogen;  
KW food supplement.  
XX  
OS Homo sapiens.  
XX  
FN M0200175068-A2.  
XX  
XX 11-OCT-2001.  
PD  
XX  
XX 22-MAR-2001; 2001MO-US009369.  
PF  
XX 30-MAR-2000; 2000US-00539330.  
PR 04-DEC-2000; 2000US-00729674.  
XX  
XX (GENM ) GENETICS INST INC.  
PA  
XX  
PI Jacobs K, Mccoy JM, Lavallie E, Collins-Racie LA, Evans C,  
PI Treacy M, Agostino M, Steininger RJ, Spaulding V, Wong GG, Clark H,  
PI Fechtel K, Merberg D;  
DR WPI: 2001-639363/73.  
DR P-PSDB; AAU39075.  
XX  
PT Secreted human proteins, useful as vaccine for treating various diseases  
PT such as autoimmune disorders (e.g. multiple sclerosis), and nervous  
PT system disorders (e.g. stroke).  
XX  
PS Disclosure, Page 580; 619pp; English.  
XX  
XX The invention relates to novel human secreted proteins, the nucleic acids  
CC encoding them. The protein may exhibit cytokine, cell proliferation or  
CC cell differentiation activity or may induce production of other cytokines  
CC in certain cell populations and may exhibit immune stimulating or immune  
CC suppressing activity, which is useful for the treatment of various immune  
CC deficiencies and disorders e.g. severe combined immunodeficiency (SCID),  
CC autoimmune disorders e.g. multiple sclerosis, systemic lupus  
CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation.  
CC The proteins are also useful in the treatment of diseases and disorders  
CC including tissue, skin and organ transplantation and in graft-versus-host  
CC diseases (GvHD), in the induction of tumour immunity, myeloid or lymphoid  
CC cell deficiencies, wound healing and tissue repair, in the treatment of  
CC burns, incisions and ulcers; as well as in treatment of periodontal  
CC disease, osteoporosis or osteoarthritis, mediated by inflammatory  
CC processes, diseases of the peripheral nervous system, Alzheimer's,  
CC Parkinson's disease, Huntington's disease, amyotrophic lateral  
CC sclerosis, and Shy-Drager syndrome, infections, infarction of cardiac and  
CC central nervous system vessel e.g. stroke, sepsis, inflammatory bowel  
CC disease, ulcers, bone regeneration. The protein, having activin- or  
CC inhibin-related activities is useful as a contraceptive based on the  
CC ability of inhibins to decrease fertility in female mammals and decrease  
CC spermatogenesis in male mammals. The proteins and nucleic acids are also  
CC useful as food supplements. The present sequence encodes a secreted  
CC protein of the invention  
XX  
SQ Sequence 1534 BP; 495 A; 363 C; 345 G; 331 T; 0 U; 0 Other;

Query Match 27.8%; Score 41.2; DB 4; Length 1534;  
Best Local Similarity 57.9%; Pred. No. 0.00022;  
Matches 73; Conservative 0; Mismatches 53; Indels 0; Gaps 0

2 CCGCTGATGAAACCGGAGAAAGTTCTTTTACCAATTCAGAAATGTCGCGTGGGCTTAAGCT 61  
633 CACTCGATGGATCAACCCACATTCACCTTTCACATTTCAATGAACCTTGGGTGAGGA 6922  
62 CCGCGTGAAGACTTAAGCTGTGCTACGTAGTGAAGAGCGGTGACAGTGTCAATCTTTTCA 121  
Db 693 CCGGATGAGACTTAAGCTGTGCTTATAGAGTGAAGCGGATGACATGATGACCTGGGTCTG 752  
122 CTGGAC 127

Db	753	CTGAC 758
RESULT 15		
ABA90962		
ID	ABA90962	standard; cDNA, 1534 BP.
XX	AC	ABA90962;
XX	DT	14-FEB-2002 (first entry)
XX	DE	Human polynucleotide SEQ ID NO 173.
XX	XX	Human, clone bcd06-7; clone yb8-1; ATCC number 98599; gene therapy; immune disorder; bacterial infection; fungal infection; cancer; tumour; autoimmune disorder; systemic lupus erythematosus; wound; ulcer; inhibin osteoporosis; osteoarthritis; nervous system disorder; neuropathy; Alzheimer's disease; Parkinson's disease; Huntington's disease; activin; haemophilia; cardiac infarction; stroke; sepsis; arthritis; vulnery; ischaemia-reperfusion injury; inflammatory bowel disease; chemotactic; Crohn's disease; cytostatic; anti-inflammatory; immunomodulator; neuroprotective; haemostatic; thrombolytic; anti-inflammatory; ss.
XX	OS	Homo sapiens.
XX	XX	US2001039335-A1.
XX	PD	08-NOV-2001.
XX	PF	04-DEC-2000; 2000US-00729674.
XX	PR	26-NOV-1997; 97US-0126425P.
XX	PR	04-DEC-1997; 97US-0067454P.
XX	PR	20-DEC-1997; 97US-0068379P.
XX	PR	02-JAN-1998; 98US-0070346P.
XX	PR	07-JAN-1998; 98US-0070643P.
XX	PR	08-JAN-1998; 98US-0070755P.
XX	PR	13-JAN-1998; 98US-0071034P.
XX	PR	22-JAN-1998; 98US-0072134P.
XX	PR	30-JAN-1998; 98US-0073095P.
XX	PR	18-FEB-1998; 98US-0075038P.
XX	PR	23-NOV-1998; 98US-00197886.
XX	XX	30-MAR-2000; 2000US-00535330.
XX	PA	(JACO/) JACOBS K.
XX	PA	(MCCO/) MCCOY J M.
XX	PA	(LAVA/) LAVALLIE E R.
XX	PA	(COLL/) COLLINS-RACIE L A.
XX	PA	(EVAN/) EVANS C.
XX	PA	(MERB/) MERBERG D.
XX	PA	(TREA/) TREACY M.
XX	PA	(AGOS/) AGOSTINO M J.
XX	PA	(STEI/) STEININGER R J.
XX	PA	(SPAU/) SPAULDING V.
XX	PA	(WONG/) WONG G G.
XX	PA	(CLAR/) CLARK H.
XX	PA	(FECH/) FECHTEL K.
XX	XX	Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C, Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V, Wong GG, Clark H, Fechtel K;
XX	DR	WPI: 2002-040725/05.
XX	XX	P-PSDB; ABB55784.
XX	PT	New secreted proteins and encoding polynucleotides, useful in gene therapies, particularly for preventing or treating autoimmune disorders, cancer, graft-versus-host disease, wound, osteoporosis, stroke or inflammations.
XX	XX	Disclosure; Page 310-311, 349pp; English.



CC The invention relates to isolated polynucleotides (ABA90876-ABA90968 and  
 CC ABA90980) and encoded proteins (ABBS5698-ABBS5800), especially  
 CC polynucleotides SEQ ID NO 1 (ABA90876) and SEQ ID NO 19 (ABA90885) and  
 CC proteins SEQ ID NO 2 (ABBS5698) and SEQ ID NO 20 (ABBS5707) contained in  
 CC clones b4306-7 and yb8-1 respectively and the clones b4306-7 and yb8-1  
 CC are deposited with the American Type Culture Collection (ATCC) with  
 CC accession number 98599. The polynucleotides and encoded polypeptides have  
 CC cytosolic, anti-inflammatory, immunomodulator, vulnery,  
 CC neuroprotective, activin, inhibin, chemotactic, haemostatic, thrombolytic  
 CC and anti-inflammatory activity and acting as cytokine modulators,  
 CC haematopoiesis regulators, tissue growth modulators and/or cadherin  
 CC suppressors. The polypeptides and polynucleotides are useful in gene  
 CC therapies, particularly for preventing, treating or ameliorating any of  
 CC the following diseases: immune deficiency and disorders; e.g. bacterial  
 CC or fungal infections, autoimmune disorders, cancer, systemic lupus  
 CC erythematosus or graft-versus-host disease; myeloid or lymphoid cell  
 CC deficiencies; wound, burns, incisions and ulcers, osteoporosis or  
 CC osteoarthritis; central and peripheral nervous system diseases and  
 CC neuropathies, e.g. Alzheimer's, Parkinson's disease, Huntington's  
 CC disease; amyotrophic lateral sclerosis or Shy-Drager syndrome;  
 CC haemophilia, cardiac infarction or stroke; inflammations, shock, sepsis  
 CC or systemic inflammatory response syndrome, ischaemia-reperfusion injury,  
 CC endotoxin lethality, arthritis, inflammatory bowel disease or Crohn's  
 CC disease; or tumours or cancers, pemphigus vulgaris or pemphigus foliaceus  
 XX

SQ Sequence 1534 BP; 495 A; 363 C; 345 G; 331 T; 0 U; 0 Other;

Query Match 27.8%; Score 41.2; DB 6; Length 1534;

Best Local Similarity 57.9%; Pred. No. 0.00022;

Matches 73; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY	2	CTCTTGATGAACCGAGAGAACTTCTTACCAATTCAAAATGTCGGCTAGAGGT	61
DB	633	CACTCGATGATCCACCCACATTCACCTTCACTTAACAATGAACTTGGTCAGAGGA	692
QY	62	CGGCGTAGACCTACCTGCTAGTAGTAAGAGGCGTGACAGTGCTACATCCTTTCA	121
DB	693	CGGCATAGACTTACCTGTGTATAGAGTGAGCGCATGCACAATGACCTGGGTCTG	752
QY	122	CTGGAC	127
DB	753	CTGAAC	758

Search completed: March 12, 2004, 18:08:07  
 Job time : 61.2641 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 12, 2004, 17:39:25 / Search time 11.019 Seconds

(without alignments)  
7458.540 Million cell updates/sec

Title: US-09-966-880A-12

Perfect score: 148

Sequence: 1 cctcttgatgacgagcgagga.....ttggtatcttcgcaataag 148

Scoring table: IDENTITY\_NTC

Gapop 10.0, Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database:

Issued Patents NA:  
1: /cgn2\_6/prodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/prodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/prodata/2/ina/5A\_COMB.seq.\*  
4: /cgn2\_6/prodata/2/ina/5B\_COMB.seq.\*  
5: /cgn2\_6/prodata/2/ina/5A\_COMB.seq.\*  
6: /cgn2\_6/prodata/2/ina/5B\_COMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	31.8	21.5	12222	4	US-09-328-925-42
2	27.8	18.8	3984	4	US-09-848-294-1
3	26.8	18.1	610	1	US-08-816-241-2
4	26.8	18.1	610	3	US-09-128-395-2
5	26.6	18.0	2208	4	US-09-543-681A-3435
6	25.6	17.3	3586	3	US-08-847-296B-3
7	25.6	17.3	14561	4	US-09-392-714-1
8	25.4	17.2	8981	4	US-09-526-193A-20
9	25.4	17.2	62804	4	US-09-800-960-3
10	25.4	17.2	62804	4	US-09-419-568F-25
11	25.2	17.0	4797	4	US-09-354-23B-25
12	25.2	17.0	4797	4	US-09-354-23B-25
13	25.2	16.9	696	4	US-09-252-991A-13591
14	25.2	16.9	1077	4	US-09-252-991A-13599
15	25.2	16.9	1467	4	US-09-252-991A-13560
16	25.2	16.9	2049	4	US-09-252-991A-13850
17	24.6	16.6	670	4	US-09-231-017B-569
18	24.6	16.6	864	4	US-09-924-570A-1
19	24.6	16.6	1082	4	US-09-247-155-181
20	24.6	16.6	1394	4	US-09-247-155-176
21	24.6	16.6	1526	4	US-09-221-017B-743
22	24.6	16.6	1613	4	US-08-965-048-3
23	24.6	16.6	1817	4	US-08-965-048-1
24	24.6	16.6	1857	4	US-09-620-312D-879
25	24.6	16.6	18596	4	US-09-318-448-11
26	24.6	16.6	18597	4	US-09-966-665-8
27	24.6	16.6	18597	4	US-09-963-333-8

28	24.6	16.6	20966	4	US-09-964-880-3	Sequence 3, Appl1
29	24.6	16.6	20966	4	US-10-277-032-3	Sequence 3, Appl1
30	24.6	16.6	45716	4	US-08-965-048-5	Sequence 5, Appl1
31	24.6	16.6	45989	4	US-08-965-048-6	Sequence 6, Appl1
32	24.4	16.5	1021	4	US-09-221-017B-750	Sequence 750, App
33	24.4	16.5	16075	3	US-09-096-942-1	Sequence 1, Appl1
34	24.4	16.5	16075	3	US-09-096-867-1	Sequence 1, Appl1
35	24.4	16.5	1830121	4	US-09-557-884-1	Sequence 1, Appl1
36	24.4	16.5	1830121	4	US-09-643-990A-1	Sequence 1, Appl1
37	24.4	16.5	4403765	3	US-09-103-840A-2	Sequence 2, Appl1
38	24.4	16.5	4411529	3	US-09-103-840A-1	Sequence 1, Appl1
39	23.8	16.1	420	3	US-09-060-756-417	Sequence 417, App
40	23.8	16.1	420	4	US-09-670-314-417	Sequence 417, App
41	23.8	16.1	588	1	US-08-742-023-4	Sequence 4, Appl1
42	23.8	16.1	588	1	US-08-742-023-5	Sequence 5, Appl1
43	23.8	16.1	588	3	US-08-968-505-4	Sequence 4, Appl1
44	23.8	16.1	588	3	US-08-968-505-5	Sequence 5, Appl1
45	23.8	16.1	2076	3	US-08-123-934A-3	Sequence 3, Appl1

## ALIGNMENTS

RESULT 1  
US-09-328-925-42/c  
Sequence 42, Application US/09328925  
Patent No. 6610906  
GENERAL INFORMATION:  
APPLICANT: Kurachi, Kotoku  
APPLICANT: Kurachi, Sumio  
TITLE OF INVENTION: Nucleotide Sequences for Gene Regulation and Methods of  
FILE REFERENCE: US-03503  
CURRENT FILING DATE: 1999-06-09  
NUMBER OF SEQ ID NOS: 84  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 42  
LENGTH: 12222  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-328-925-42

Query Match 21.5%, Score 31.8, DB 4; Length 12222;  
Best Local Similarity 53.7%; Pred. No. 0.047;  
Matches 66; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 18 GGAAGTTCTTACCAATCAAAATGTCGGCTGAGGCGGCGTGAACCTTACC 77  
DB 202 GGCCTTTGTTTCCATCTGCACATGACCCCACTGAAGCCTGCTTCTCCAC 143  
QY 78 TGTGCTACGTGTGAAGAGCGGTGACAGTGTACATCTTTCACTGACCTTGTATC 137  
DB 142 TGTGCAATGTAGTACGACGACGACGACATGTTCACCTTCTCTCCCAAGCTTC 83  
QY 138 TTC 140  
DB 82 CTC 80  
RESULT 2  
US-09-848-294-1/c  
Sequence 1, Application US/09848294  
Patent No. 6479640  
GENERAL INFORMATION:  
APPLICANT: Tonks, Nicholas K.  
TITLE OF INVENTION: Isolation of A CDNA Encoding A No. 6479640el  
TITLE OF INVENTION: Protein Tyrosine Phosphatase Which Localizes to Focal  
FILE REFERENCE: CSHL90-04PZA  
CURRENT FILING DATE: 2001-05-03  
CURRENT APPLICATION NUMBER: US/09/848,294  
PRIOR APPLICATION NUMBER: 09/235,251



PRIOR FILING DATE: 1999-01-22  
PRIOR APPLICATION NUMBER: 08/759,536  
PRIOR FILING DATE: 1996-12-04  
PRIOR APPLICATION NUMBER: 08/107,420  
PRIOR FILING DATE: 1993-08-16  
PRIOR APPLICATION NUMBER: 07/663,579  
PRIOR FILING DATE: 1991-03-01  
PRIOR APPLICATION NUMBER: 07/494,036  
PRIOR FILING DATE: 1990-03-14  
NUMBER OF SEQ ID NOS: 13  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 1  
LENGTH: 3984  
TYPE: DNA  
ORGANISM: Homosapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (24)...(2765)  
US-09-848-294-1

Query Match 18.8%; Score 27.8; DB 4; Length 3984;  
Best Local Similarity 53.2%; Pred. No. 1.1;  
Matches 59; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 27 TTTACCAATTCAAAATGTCCTGCGCTAAGGTCGCGGTAGACCTACCTGTGTAGC 86  
DB 2500 TTTACCAATTCAAAATGTCGAGAGATCATCGGATACCGGTGACGCAATGCGACG 2441  
QY 87 TAGTGAAGAGCGCGTGAAGTGTACATCTTTTCACTGACCTTGGTATC 137  
DB 2440 TAGTGAAGATGTGTCACTGTGTCTTCTCCCGGTGCGGTGTTGTAGCC 2390

## RESULT 3

US-08-816-241-2  
Sequence 2, Application US/08816241  
Patent No. 5804185  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Goll, Surya K.  
TITLE OF INVENTION: NOVEL RNA EDITING ENZYME  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/816,241  
FILING DATE: Filed Herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0239 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 610 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single

TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: PROSTUT09  
CLONE: 1646833  
US-08-816-241-2

Query Match 18.1%; Score 26.8; DB 1; Length 610;  
Best Local Similarity 57.0%; Pred. No. 1;  
Matches 49; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 8 ATGAACCGAGAGAGTTCTTACCAATTCAAAATGTCCTGCGCTAAGGTCGCGCT 67  
DB 34 ATGTATCCAGGCAATTCATCTTCAATTTAAACCTATGGAAGCAACGATCGAAC 93  
QY 68 GAGACCTACCTGCTGCTAGTAGAA 93  
DB 94 GAAACTGTGCTGTCTCACCCTGGA 119

## RESULT 4

US-09-128-395-2  
Sequence 2, Application US/09128395  
Patent No. 6087108  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Goll, Surya K.  
TITLE OF INVENTION: NOVEL RNA EDITING ENZYME  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/128,395  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/816,241  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0239 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 610 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: PROSTUT09  
CLONE: 1646833  
US-09-128-395-2

Query Match 18.1%; Score 26.8; DB 3; Length 610;  
Best Local Similarity 57.0%; Pred. No. 1;  
Matches 49; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 8 ATGAACCGAGAGAGTTCTTACCAATTCAAAATGTCCTGCGCTAAGGTCGCGCT 67  
DB 34 ATGTATCCAGGCAATTCATCTTCAATTTAAACCTATGGAAGCAACGATCGAAC 93



QY 68 GAGACCTACCTGTGCTACGTAGTAA 93  
DB 94 GAACTTGGCTGTGCTTACCGTGA 119

## RESULT 5

US-09-543-681A-3435  
; Sequence 3435, Application US/09543681A  
; Patent No. 6605709  
; GENERAL INFORMATION:  
; APPLICANT: GARY BRETON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
; FILE REFERENCE: 2709 1002-001  
; CURRENT APPLICATION NUMBER: US/09/543,681A  
; PRIOR FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/128,706  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 3435  
; LENGTH: 2208  
; TYPE: DNA  
; ORGANISM: Proteus mirabilis  
US-09-543-681A-3435

Query Match 18.0%; Score 26.6; DB 4; Length 2208;  
Best Local Similarity 56.2%; Pred. No. 2.4;  
Matches 50; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 6 TGATGAACCGAGAGATTCTTACCAATTCAAATGCGCTGGCTAAGGTCGGC 65  
DB 498 TCAAAATATAGGAACTTCATTAGCAATTCAAAATATAGAGCTGTACACTTACAG 557  
QY 66 GTGAGACCTACTGTGTCTACGTAGTGAAG 94  
DB 558 GAGACTGGAGTGTGTACTTAAATAGTGAAG 586

## RESULT 6

US-08-847-296B-3  
; Sequence 3, Application US/08847296B  
; Patent No. 6271347  
; GENERAL INFORMATION:  
; APPLICANT: DAUGHERTY, BRUCE L.  
; APPLICANT: DEMARTINO, JULIE A.  
; APPLICANT: SICILIANO, SALVATORE J.  
; APPLICANT: SPRINGER, MARTIN J.  
; TITLE OF INVENTION: EOSINOPHIL EOTAXIN RECEPTOR  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Merck & Co., Inc.  
; STREET: P.O. Box 2000, 126 E. Lincoln Ave.  
; CITY: Rahway  
; STATE: NJ  
; COUNTRY: USA  
; ZIP: 07065-0900  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/847,296B  
; FILING DATE: 24-APR-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/016,158  
; FILING DATE: 26-APR-1996  
; APPLICATION NUMBER: 60/017,113  
; FILING DATE: 26-APR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Thies, J. Eric  
; REGISTRATION NUMBER: 35,382

REFERENCE/DOCKET NUMBER: 19634Y  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 908-594-3904  
; TELEFAX: 908-594-4720  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3586 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
US-08-847-296B-3

Query Match 17.3%; Score 25.6; DB 3; Length 3586;  
Best Local Similarity 50.8%; Pred. No. 7.4;  
Matches 61; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 21 AGTTCTTACCAATTCAAAATGTCGCTGAGGCTGAGAGCTTACCTGT 80  
DB 1893 AGTCATTTCCTCTCTCAAAAGCCTGGAATGTGCTATTGATGAGAGATTAAGTCTGG 1952  
QY 81 GCTACGTGTAAGAGGCGTGACAGTGTCTACATCTTTTCACTGGAAGTTGGTTATCTTC 140  
DB 1953 AAGACCCATGATAAAGAGATCAACAAGTCCACCAGGAGCCTTATTTTCTTAATTTTC 2012

## RESULT 7

US-09-392-714-1/c  
; Sequence 1, Application US/09392714A  
; Patent No. 6666147  
; GENERAL INFORMATION:  
; APPLICANT: Scanlan, Matthew J.  
; APPLICANT: Gure, Ali O.  
; APPLICANT: Williamson, Barbara  
; APPLICANT: Chen, Yao-Tseng  
; APPLICANT: Old, Lloyd J.  
; TITLE OF INVENTION: Cancer Associated Antigens and Uses  
; FILE REFERENCE: 10461/7062  
; CURRENT APPLICATION NUMBER: US/09/392,714A  
; PRIOR FILING DATE: 1999-09-09  
; EARLIER APPLICATION NUMBER: PCT/US98/14679  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 14561  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-392-714-1

Query Match 17.3%; Score 25.6; DB 4; Length 14561;  
Best Local Similarity 55.7%; Pred. No. 15;  
Matches 49; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 39 AAAATGTCGCTGGGCTTAAGGCTGAGACTGCTGTGCTAGTGAAGAGGC 98  
DB 9032 AAGACGACGAGAGAGAGAGAGAGAGCTGGAATGTGAGCTGAGACTGGAAGCGCTAAGGC 8973  
QY 99 GTGACAGTGTACATCTTTTCACTGGA 126  
DB 8972 GTGACACTGTACTTGTGTGCAAGGA 8945

## RESULT 8

US-09-526-193A-20/c  
; Sequence 20, Application US/09526193A  
; Patent No. 6617122  
; GENERAL INFORMATION:  
; APPLICANT: Hayden, Michael R.  
; APPLICANT: Brooks-Wilson, Angela R.  
; APPLICANT: Pimstone, Simon N.



;; TITLE OF INVENTION: METHODS AND REAGENTS FOR MODULATING  
;; TITLE OF INVENTION: CHOLESTEROL LEVELS  
;; FILE REFERENCE: 50110/002005  
;; CURRENT APPLICATION NUMBER: US/09/526,193A  
;; CURRENT FILING DATE: 2000-03-15  
;; PRIOR APPLICATION NUMBER: 60/124,702  
;; PRIOR FILING DATE: 1999-03-15  
;; PRIOR APPLICATION NUMBER: 60/138,048  
;; PRIOR FILING DATE: 1999-06-08  
;; PRIOR APPLICATION NUMBER: 60/139,600  
;; PRIOR FILING DATE: 1999-06-17  
;; PRIOR APPLICATION NUMBER: 60/151,977  
;; PRIOR FILING DATE: 1999-09-01  
;; NUMBER OF SEQ ID NOS: 287  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 20  
;; LENGTH: 8981  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
US-09-526-193A-20

Query Match 17.2%; Score 25.4; DB 4; Length 8981;  
Best Local Similarity 61.2%; Pred. No. 14;  
Matches 41; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 44 GTCCGCTGGGTAGAGGTGGCGGTGACCTACTCTGCTACGTAGTGAAGGCGTGC 103  
DB 7558 GTCTGGTGGGCAATGGACATGTAGACATCTAACGCTCTACAGAGGAGATATACAC 7599

QY 104 AGTGCTA 110  
DB 7598 AGGCCAA 7592

RESULT 9  
US-09-800-960-3  
;; Sequence 3, Application US/09800960  
;; Patent No. 6387677  
;; GENERAL INFORMATION:  
;; APPLICANT: YE, Jane et al.  
;; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
;; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
;; TITLE OF INVENTION: THEREOF  
;; FILE REFERENCE: CI001158  
;; CURRENT APPLICATION NUMBER: US/09/800,960  
;; CURRENT FILING DATE: 2001-03-08  
;; NUMBER OF SEQ ID NOS: 4  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 3  
;; LENGTH: 62804  
;; TYPE: DNA  
;; ORGANISM: Human  
;; FEATURE:  
;; NAME/KEY: misc\_feature  
;; LOCATION: (1)...(62804)  
;; OTHER INFORMATION: n = A,T,C or G  
US-09-800-960-3

Query Match 17.2%; Score 25.4; DB 4; Length 62804;  
Best Local Similarity 58.7%; Pred. No. 36;  
Matches 44; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 19 GAAGTTCTTACCAATCAAAATGTCGCTGAGGTGAGGTGAGACCTACTCT 78  
DB 24634 GAATTTCTTAGGATTCGCAATGTCTCCCTGCGGTGGGGGGGGGGGCAATTC 24693

QY 79 GTGCTACGTAGTGA 93  
DB 24694 ATCCCACTTGA 24708

RESULT 10  
US-10-096-960-3

;; Sequence 3, Application US/10096960  
;; Patent No. 6664085  
;; GENERAL INFORMATION:  
;; APPLICANT: YE, Jane et al.  
;; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
;; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
;; TITLE OF INVENTION: THEREOF  
;; FILE REFERENCE: CI001158D1V  
;; CURRENT APPLICATION NUMBER: US/10/096,960  
;; CURRENT FILING DATE: 2002-03-14  
;; PRIOR APPLICATION NUMBER: 09/800,960  
;; PRIOR FILING DATE: 2001-03-08  
;; NUMBER OF SEQ ID NOS: 4  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 3  
;; LENGTH: 62804  
;; TYPE: DNA  
;; ORGANISM: Homo sapien  
;; FEATURE:  
;; NAME/KEY: misc\_feature  
;; LOCATION: (1)...(62804)  
;; OTHER INFORMATION: n = A,T,C or G  
US-10-096-960-3

Query Match 17.2%; Score 25.4; DB 4; Length 62804;  
Best Local Similarity 58.7%; Pred. No. 36;  
Matches 44; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 19 GAAGTTCTTACCAATCAAAATGTCGCTGAGGTGAGGTGAGACCTACTCT 78  
DB 24634 GAATTTCTTAGGATTCGCAATGTCTCCCTGCGGTGGGGGGGGGGGCAATTC 24693

QY 79 GTGCTACGTAGTGA 93  
DB 24694 ATCCCACTTGA 24708

RESULT 11  
US-09-419-568F-25  
;; Sequence 25, Application US/09419568F  
;; Patent No. 6331613  
;; GENERAL INFORMATION:  
;; APPLICANT: Dumoutier, Laure  
;; APPLICANT: Renauld, Jean-Christophe  
;; TITLE OF INVENTION: Isolated Nucleic Acid Molecules which Encode T Cell Inducible Fa  
;; TITLE OF INVENTION: (Tifs) The Proteins Encoded, and Uses Thereof  
;; FILE REFERENCE: LUD 5543.2  
;; CURRENT APPLICATION NUMBER: US/09/419,568F  
;; CURRENT FILING DATE: 1999-10-18  
;; PRIOR APPLICATION NUMBER: US09/354,243  
;; PRIOR FILING DATE: 1999-07-16  
;; PRIOR APPLICATION NUMBER: US09/178,973  
;; PRIOR FILING DATE: 1998-10-26  
;; NUMBER OF SEQ ID NOS: 29  
;; SEQ ID NO 25  
;; LENGTH: 4797  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
US-09-419-568F-25

Query Match 17.0%; Score 25.2; DB 4; Length 4797;  
Best Local Similarity 66.7%; Pred. No. 12;  
Matches 36; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 87 TAGTGAAGAGGCGGTACAGCTGTACTTCTTCACTGGAATTTGGTATCTTC 140  
DB 3750 TAGATTAGGAATTCAGTAGGGAATGCTTTTCACTTGAATTTGGGTTCTTC 3803

RESULT 12  
US-09-354-243B-25



; Sequence 25, Application US/09354243B  
; Patent No. 6359117  
; GENERAL INFORMATION:  
; APPLICANT: Dumontier, Laure  
; APPLICANT: Louned, Jamila  
; APPLICANT: Renaud, Jean-Christophe  
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules which Encode T Cell Inducible Fa  
; TITLE OF INVENTION: (Tifs)  
; FILE REFERENCE: LUD 5543.1  
; CURRENT APPLICATION NUMBER: US/09/354,243B  
; PRIOR FILING DATE: 1999-07-16  
; PRIOR APPLICATION NUMBER: US09/178,973  
; PRIOR FILING DATE: 1998-10-26  
; NUMBER OF SEQ ID NOS: 29  
; SEQ ID NO 25  
; LENGTH: 4797  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-354-243B-25

Query Match 17.0%; Score 25.2; DB 4; Length 4797;  
Best Local Similarity 66.7%; Pred. No. 12;  
Matches 36; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 87 TAGTGAAGGCGGTGACATGCTTACATCTTTTCACTGAGACTTGGTATCTTC 140  
DB 3750 TAGATTAGAGAAATTTCAGTAGGAATGCTTTTCACTGAAATTGGGTTCTTC 3803

RESULT 13  
US-09-252-991A-13391  
; Sequence 13391, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 13391  
; LENGTH: 696  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-13391

Query Match 16.9%; Score 25; DB 4; Length 696;  
Best Local Similarity 49.6%; Pred. No. 5.8;  
Matches 64; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 1 CCTCTGATGACCGGAGAGATTCTTTACCAATTCAAATATGTCGGCTGGGCTAAGG 60  
DB 393 CCGTGTCTGATGCTGGCTGGCTGTTGACCTGCAATCGCCGACGCGCCCAAGT 452  
QY 61 TCGGCGTGAAGACCTTACTGCTAGCTAGTGAAGAGCGGTGACATGCTTACATCTTTTC 120  
DB 453 GTTGCTGAGAGCGCTTCTGAGCGTTGCGCGGAGCGCGCGGAGCGCGGCGCTTGT 512  
QY 121 ACTGGAATT 129  
DB 513 CTTGGCGTT 521

RESULT 14  
US-09-252-991A-13599  
; Sequence 13599, Application US/09252991A

; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 13599  
; LENGTH: 1077  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-13599

Query Match 16.9%; Score 25; DB 4; Length 1077;  
Best Local Similarity 49.6%; Pred. No. 7.1;  
Matches 64; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 1 CCTCTGATGACCGGAGAGATTCTTTACCAATTCAAATATGTCGGCTGGGCTAAGG 60  
DB 31 CCGTGTCTGATGCTGGCTGGCTGTTGACCTGCAATCGCCGACGCGCCCAAGT 90  
QY 61 TCGGCGTGAAGACCTTACTGCTAGCTAGTGAAGAGCGGTGACATGCTTCTTTTC 120  
DB 91 GTTGCTGAGAGCGCTTCTGAGCGTTGCGCGGAGCGCGCGGAGCGCGGCGCTTGT 150

QY 121 ACTGGAATT 129  
DB 151 CTTGGCGTT 159

RESULT 15  
US-09-252-991A-13960/C  
; Sequence 13960, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 13960  
; LENGTH: 1467  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-13960

Query Match 16.9%; Score 25; DB 4; Length 1467;  
Best Local Similarity 49.6%; Pred. No. 8.3;  
Matches 64; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 1 CCTCTGATGACCGGAGAGATTCTTTACCAATTCAAATATGTCGGCTGGGCTAAGG 60  
DB 1244 CCGTGTCTGATGCTGGCTGGCTGTTGACCTGCAATCGCCGACGCGCCCAAGT 1185  
QY 61 TCGGCGTGAAGACCTTACTGCTAGCTAGTGAAGAGCGGTGACATGCTTCTTTTC 120  
DB 1184 GTTGCTGAGAGCGCTTCTGAGCGTTGCGCGGAGCGCGCGGAGCGCGGCGCTTGT 1125  
QY 121 ACTGGAATT 129  
DB 1124 CTTGGCGTT 1116



Mon Mar 15 09:27:54 2004

us-09-966-880a-12.rn1

Page 6

Search completed: March 13, 2004, 00:51:14  
Job time : 19.0119 secs

---







QY 121 ACTGACTTGGTTATCTTCCCAATAAG 148  
DB 121 ACTGACTTGGTTATCTTCCCAATAAG 148

## RESULT 2

US-09-966-880A-7  
; Sequence 7, Application US/09966880A  
; Patent No. US20020164743A1  
; GENERAL INFORMATION:  
; APPLICANT: Honjo, Tasuku  
; APPLICANT: Muramatsu, Masamichi  
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
; FILE REFERENCE: 06501-088001  
; CURRENT APPLICATION NUMBER: US/09/966,880A  
; PRIOR FILING DATE: 2001-09-28  
; PRIOR APPLICATION NUMBER: PCT/JP00/01918  
; PRIOR FILING DATE: 2000-03-28  
; PRIOR APPLICATION NUMBER: JP 11-371382  
; PRIOR FILING DATE: 1999-12-27  
; PRIOR APPLICATION NUMBER: JP 11-178999  
; PRIOR FILING DATE: 1999-06-24  
; PRIOR APPLICATION NUMBER: JP 11-87192  
; PRIOR FILING DATE: 1999-03-29  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7  
; LENGTH: 2818  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; NAME/KEY: CDS  
; LOCATION: (80)...(673)  
; FEATURE:  
; NAME/KEY: 5'UTR  
; LOCATION: (1)...(79)  
; FEATURE:  
; NAME/KEY: 3'UTR  
; LOCATION: (677)...(2818)  
US-09-966-880A-7

Query Match 100.0%; Score 148; DB 9; Length 2818;  
Best Local Similarity 100.0%; Pred. No. 1.8e-45;  
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTTTGATGACCGGAGGAAGTTCTTTACCAATTCAAAATGTCCGCTGGGCTAAAGG 60  
DB 88 CCTTTGATGACCGGAGGAAGTTCTTTACCAATTCAAAATGTCCGCTGGGCTAAAGG 147  
QY 61 TCGGCGTAGACCTACCTGCTGCTACGTAGTGAAGAGCGGTGACAGTCTACATCTTTTC 120  
DB 148 TCGGCGTAGACCTACCTGCTGCTACGTAGTGAAGAGCGGTGACAGTCTACATCTTTTC 207  
QY 121 ACTGACTTGGTTATCTTCCCAATAAG 148  
DB 208 ACTGACTTGGTTATCTTCCCAATAAG 235

## RESULT 3

US-09-966-880A-10  
; Sequence 10, Application US/09966880A  
; Patent No. US20020164743A1  
; GENERAL INFORMATION:  
; APPLICANT: Honjo, Tasuku  
; APPLICANT: Muramatsu, Masamichi  
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
; FILE REFERENCE: 06501-088001  
; CURRENT APPLICATION NUMBER: US/09/966,880A  
; PRIOR FILING DATE: 2001-09-28  
; PRIOR APPLICATION NUMBER: PCT/JP00/01918  
; PRIOR FILING DATE: 2000-03-28  
; PRIOR APPLICATION NUMBER: JP 11-371382

; PRIOR FILING DATE: 1999-12-27  
; PRIOR APPLICATION NUMBER: JP 11-178999  
; PRIOR FILING DATE: 1999-06-24  
; PRIOR APPLICATION NUMBER: JP 11-87192  
; PRIOR FILING DATE: 1999-03-29  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10  
; LENGTH: 6564  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-966-880A-10  
Query Match 100.0%; Score 148; DB 9; Length 6564;  
Best Local Similarity 100.0%; Pred. No. 2.6e-45;  
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTTTGATGACCGGAGGAAGTTCTTTACCAATTCAAAATGTCCGCTGGGCTAAAGG 60  
DB 1065 CCTTTGATGACCGGAGGAAGTTCTTTACCAATTCAAAATGTCCGCTGGGCTAAAGG 1124  
QY 61 TCGGCGTAGACCTACCTGCTGCTACGTAGTGAAGAGCGGTGACAGTCTACATCTTTTC 120  
DB 1125 TCGGCGTAGACCTACCTGCTGCTACGTAGTGAAGAGCGGTGACAGTCTACATCTTTTC 1184  
QY 121 ACTGACTTGGTTATCTTCCCAATAAG 148  
DB 1185 ACTGACTTGGTTATCTTCCCAATAAG 1212

## RESULT 4

US-09-966-880A-35  
; Sequence 35, Application US/09966880A  
; Patent No. US20020164743A1  
; GENERAL INFORMATION:  
; APPLICANT: Honjo, Tasuku  
; APPLICANT: Muramatsu, Masamichi  
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
; FILE REFERENCE: 06501-088001  
; CURRENT APPLICATION NUMBER: US/09/966,880A  
; PRIOR FILING DATE: 2001-09-28  
; PRIOR APPLICATION NUMBER: PCT/JP00/01918  
; PRIOR FILING DATE: 2000-03-28  
; PRIOR APPLICATION NUMBER: JP 11-371382  
; PRIOR FILING DATE: 1999-12-27  
; PRIOR APPLICATION NUMBER: JP 11-178999  
; PRIOR FILING DATE: 1999-06-24  
; PRIOR APPLICATION NUMBER: JP 11-87192  
; PRIOR FILING DATE: 1999-03-29  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 35  
; LENGTH: 11204  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-966-880A-35

Query Match 100.0%; Score 148; DB 9; Length 11204;  
Best Local Similarity 100.0%; Pred. No. 3.3e-45;  
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTTTGATGACCGGAGGAAGTTCTTTACCAATTCAAAATGTCCGCTGGGCTAAAGG 60  
DB 6280 CCTTTGATGACCGGAGGAAGTTCTTTACCAATTCAAAATGTCCGCTGGGCTAAAGG 6339  
QY 61 TCGGCGTAGACCTACCTGCTGCTACGTAGTGAAGAGCGGTGACAGTCTACATCTTTTC 120  
DB 6340 TCGGCGTAGACCTACCTGCTGCTACGTAGTGAAGAGCGGTGACAGTCTACATCTTTTC 6399  
QY 121 ACTGACTTGGTTATCTTCCCAATAAG 148  
DB 6400 ACTGACTTGGTTATCTTCCCAATAAG 6427



```
RESULT 5
US-09-966-880a-1
; Sequence 1, Application US/09966880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tsauku
; APPLICANT: Muramatsu, Masamichi
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966,880A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2440
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (93)...(686)
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1)...(92)
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: (690)...(2440)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(2440)
; OTHER INFORMATION: n = A,T,C or G
US-09-966-880a-1

Query Match
Best Local Similarity 74.1%; Score 109.6; DB 9; Length 2440;
; Pred. No. 7.2e-31;
Matches 124; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 1 CCTTTATGATGACCGAGAGATTCTTTACCAATTCAAAATGTCGCTGGCTAAGG 60
DB 101 CCTTCTGATGAGCAAGAAAGATTCTTTACCAATTCAAAATGTCGCTGGCTAAGG 160
QY 61 TCGGCGTGAAGACCTGCTGCTAGCTAGTGAAGAGCGGTGACGTACATCTTTTC 120
DB 161 ACCGCGATGAGACCTGCTGCTAGCTGCTGTAAGAGAGAGAGATGTCACCTCTGCTC 220
QY 121 ACTGACTTGGTTATCTTTCGCAATAG 148
DB 221 ACTGACTTGGCGCACTTTCGCAACAG 248

RESULT 6
US-09-867-701-3576
; Sequence 3576, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.457
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 3576
; LENGTH: 254
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(254)
; OTHER INFORMATION: n = A,T,C or G
US-09-867-701-3576

Query Match
Best Local Similarity 27.8%; Score 41.2; DB 9; Length 254;
; Pred. No. 3e-05;
Matches 73; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 2 CTTTGTATGACCGAGAGATTCTTTACCAATTCAAAATGTCGCTGGCTAAGGT 61
DB 7 CACTGATGATCCACCACCATTCATCTTCAACATGAACTTGGTCAGAGGA 66
QY 62 CGGCGTGAACCTGCTGCTAGCTAGTGAAGAGCGGTGACGTACATCCTTTCA 121
DB 67 CGGATGAGACTTCTGTTATGAGGTGAGCGCATGACATGACATGACCTGGCTCTG 126
QY 122 CTGGAC 127
DB 127 CTGAAC 132

RESULT 7
US-09-796-692-4007
; Sequence 4007, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/166,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4007
; LENGTH: 466
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-796-692-4007

Query Match
Best Local Similarity 27.8%; Score 41.2; DB 9; Length 466;
```



Best Local Similarity 57.9%; Pred. No. 4e-05;  
Matches 73; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 2 CTCTTGATGAACCGAGAGAGTTCTTTTACCAATTCAAAAATGTCGCTGGCTAAGGT 61  
DB 270 CACTCGATGATGATCCACCATTCCTTCACTTTAAACAATGAACTTGGGTGAGAGGA 329  
QY 62 CGGCGTGAACCTTACTGTGCTACTAGTGAAGGCGTGACAGTCTTCACTTTTCA 121  
DB 330 CGGCATGAGACTTACTGTGTATGAGTGAGCGCATGACAAATGACACTGGGCTCTG 389  
QY 122 CTGGAC 127  
DB 390 CTGAAC 395

## RESULT 8

US-10-040-862-4007  
; Sequence 4007, Application US/10040862  
; Publication No. US20030078396A1  
; GENERAL INFORMATION:  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Mannion, Jane A.  
; APPLICANT: Mannion, Jane  
; APPLICANT: Retter, Marc  
; APPLICANT: Corixa Corporation  
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy  
; TITLE OF INVENTION: Hematological Malignancies  
; FILE REFERENCE: 014058-013520US  
; CURRENT APPLICATION NUMBER: US/10/040,862  
; CURRENT FILING DATE: 2001-11-06  
; PRIOR APPLICATION NUMBER: US 60/186,126  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: US 60/190,479  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: US 60/200,545  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: US 60/200,303  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 60/200,779  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 60/200,999  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: US 60/202,084  
; PRIOR FILING DATE: 2000-05-04  
; PRIOR APPLICATION NUMBER: US 60/206,201  
; PRIOR FILING DATE: 2000-05-22  
; PRIOR APPLICATION NUMBER: US 60/218,950  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: US 60/222,903  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: US 60/223,416  
; PRIOR FILING DATE: 2000-08-04  
; PRIOR APPLICATION NUMBER: US 60/223,378  
; PRIOR FILING DATE: 2000-08-07  
; PRIOR APPLICATION NUMBER: US 09/796,692  
; PRIOR FILING DATE: 2001-03-01  
; NUMBER OF SEQ ID NOS: 10467  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 4007  
; LENGTH: 466  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-040-862-4007

Query Match 27.8%; Score 41.2; DB 14; Length 466;  
Best Local Similarity 57.9%; Pred. No. 4e-05;  
Matches 73; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 2 CTCTTGATGAACCGAGAGAGTTCTTTTACCAATTCAAAAATGTCGCTGGCTAAGGT 61  
DB 270 CACTCGATGATGATCCACCATTCCTTCACTTTAAACAATGAACTTGGGTGAGAGGA 329

QY 62 CGGCGTGAACCTTACTGTGCTACTAGTGAAGGCGTGACAGTCTTCACTTTTCA 121  
DB 330 CGGCATGAGACTTACTGTGTATGAGTGAGCGCATGACAAATGACACTGGGCTCTG 389

QY 122 CTGGAC 127  
DB 390 CTGAAC 395

## RESULT 9

US-10-057-475B-4007  
; Sequence 4007, Application US/10057475B  
; Publication No. US20040002068A1  
; GENERAL INFORMATION:  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Mannion, Jane  
; APPLICANT: Clapper, Jonathan David  
; APPLICANT: Wang, Aijun  
; APPLICANT: Ordenez, Nadia  
; APPLICANT: Carter, Lauren  
; APPLICANT: McNeill, Patricia Dianne  
; APPLICANT: Corixa Corporation  
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy  
; TITLE OF INVENTION: Hematological Malignancies  
; FILE REFERENCE: 014058-014402US  
; CURRENT APPLICATION NUMBER: US/10/057,475B  
; CURRENT FILING DATE: 2002-01-22  
; PRIOR APPLICATION NUMBER: US 60/186,126  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: US 60/190,479  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: US 60/200,545  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: US 60/200,303  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 60/200,779  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 60/200,999  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: US 60/202,084  
; PRIOR FILING DATE: 2000-05-04  
; PRIOR APPLICATION NUMBER: US 60/206,201  
; PRIOR FILING DATE: 2000-05-22  
; PRIOR APPLICATION NUMBER: US 60/218,950  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: US 60/222,903  
; PRIOR FILING DATE: 2000-08-03  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 10979  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 4007  
; LENGTH: 466  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-057-475B-4007

Query Match 27.8%; Score 41.2; DB 15; Length 466;  
Best Local Similarity 57.9%; Pred. No. 4e-05;  
Matches 73; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 2 CTCTTGATGAACCGAGAGAGTTCTTTTACCAATTCAAAAATGTCGCTGGCTAAGGT 61  
DB 270 CACTCGATGATGATCCACCATTCCTTCACTTTAAACAATGAACTTGGGTGAGAGGA 329  
QY 62 CGGCGTGAACCTTACTGTGCTACTAGTGAAGGCGTGACAGTCTTCACTTTTCA 121  
DB 330 CGGCATGAGACTTACTGTGTATGAGTGAGCGCATGACAAATGACACTGGGCTCTG 389

QY 122 CTGGAC 127  
DB 390 CTGAAC 395



```
RESULT 10
US-10-154-884B-4007
; Sequence 4007, Application US/10154884B
; Publication No. US20040005561A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc W.
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; FILE REFERENCE: 014058-013521US
; CURRENT APPLICATION NUMBER: US/10/154,884B
; CURRENT FILING DATE: 2002-05-23
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 11290
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4007
; LENGTH: 466
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-154-884B-4007

Query Match      27.8%; Score 41.2; DB 15; Length 466;
Best Local Similarity 57.9%; Pred. No. 4e-05;
Matches 73; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY      2  CTCTTGATGAACCGAGAGAGTTCTTTACCAATTCAAAATGTCGCGTGGGCTAAGGCT 61
DB      270  CACTCGATGATGCCACCACTTCCTTCAACTTAAACATGAACCTTGGGTCAAGGA 329

QY      62  CGCGGTAGACCTACTGCTGTACGTAGTGAAGAGCGGTGACAGTGTCTAATCCTTTCA 121
DB      330  CGGATGAGACTTACTGTGTATAGAGTGGAGCGCATGACCAATGACACTGGGTCTCTG 389

QY      122  CTGAGC 127
DB      390  CTGAAC 395

RESULT 11
US-09-796-692-4443
; Sequence 4443, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc W.
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; FILE REFERENCE: 2077.001200
```

```
; CURRENT APPLICATION NUMBER: US/09/796,692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4443
; LENGTH: 505
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-796-692-4443

Query Match      27.8%; Score 41.2; DB 9; Length 505;
Best Local Similarity 57.9%; Pred. No. 4.1e-05;
Matches 73; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY      2  CTCTTGATGAACCGAGAGAGTTCTTTACCAATTCAAAATGTCGCGTGGGCTAAGGCT 61
DB      270  CACTCGATGATGCCACCACTTCCTTCAACTTAAACATGAACCTTGGGTCAAGGA 329

QY      62  CGCGGTAGACCTACTGCTGTACGTAGTGAAGAGCGGTGACAGTGTCTAATCCTTTCA 121
DB      330  CGGATGAGACTTACTGTGTATAGAGTGGAGCGCATGACCAATGACACTGGGTCTCTG 389

QY      122  CTGAGC 127
DB      390  CTGAAC 395

RESULT 12
US-10-040-862-4443
; Sequence 4443, Application US/10040862
; Publication No. US20030078396A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc W.
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; FILE REFERENCE: 014058-013520US
; CURRENT APPLICATION NUMBER: US/10/040,862
; CURRENT FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
```



```

; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: US 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/223,378
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: US 09/796,692
; PRIOR FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 10467
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4443
; LENGTH: 505
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-040-862-4443
```

```

Query Match      27.8%; Score 41.2; DB 14; Length 505;
Best Local Similarity 57.9%; Pred. No. 4.1e-05;
Matches 73; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
```

```

QY 2 CTCTGATGAACCGAGAGAGTTCTTACCAATTCAAAATGTCGCTGGCTAAGGT 61
    |||||
DB 270 CACTCGATGATCCACCCACATTCCTTCACTTAACATGAACCTGGGCAAGGA 329
    |||||
QY 62 CGGCGTAGAGACTTACTGTGCTACGTAGTGAAGAGGCGTAGTGCATCCTTTTCA 121
    |||||
DB 330 CGGATAGAGACTTACTGTGTTATGAGGTGAGGCGCATGACATGACCTGGGTCCTG 389
    |||||
QY 122 CTGAGC 127
    |||||
DB 390 CTGAAC 395
```

```

RESULT 13
US-10-057-475B-4443
; Sequence 4443, Application US/10057475B
; Publication No. US20040002068A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Clapper, Jonathan David
; APPLICANT: Wang, Aijun
; APPLICANT: Ordonez, Nadia
; APPLICANT: Carter, Lauren
; APPLICANT: McNeill, Patricia Dianne
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; TITLE OF INVENTION: Hematological Malignancies
; FILE REFERENCE: 014058-01440205
; CURRENT APPLICATION NUMBER: US/10/057,475B
; CURRENT FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
```

```

; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 10979
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4443
; LENGTH: 505
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-057-475B-4443
```

```

Query Match      27.8%; Score 41.2; DB 15; Length 505;
Best Local Similarity 57.9%; Pred. No. 4.1e-05;
Matches 73; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
```

```

QY 2 CTCTGATGAACCGAGAGAGTTCTTACCAATTCAAAATGTCGCTGGCTAAGGT 61
    |||||
DB 270 CACTCGATGATCCACCCACATTCCTTCACTTAACATGAACCTGGGCAAGGA 329
    |||||
QY 62 CGGCGTAGAGACTTACTGTGCTACGTAGTGAAGAGGCGTAGTGCATCCTTTTCA 121
    |||||
DB 330 CGGATAGAGACTTACTGTGTTATGAGGTGAGGCGCATGACATGACCTGGGTCCTG 389
    |||||
QY 122 CTGAGC 127
    |||||
DB 390 CTGAAC 395
```

```

RESULT 14
US-10-154-884B-4443
; Sequence 4443, Application US/10154884B
; Publication No. US20040005561A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc W.
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therap
; TITLE OF INVENTION: Hematological Malignancies
; FILE REFERENCE: 014058-01352105
; CURRENT APPLICATION NUMBER: US/10/154,884B
; CURRENT FILING DATE: 2002-05-23
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 11290
```



SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 4443  
LENGTH: 505  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-154-884B-4443

Query Match 27.8%; Score 41.2; DB 15; Length 505;  
Best Local Similarity 57.9%; Pred. No. 4.1e-05;  
Matches 73; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 2 CTTTGATGAACCGAGAGAGTTCTTTACCAATTCAAAAATGTCGCTGCGCTAAGGGT 61  
DB 270 CACTCGATGATCCACCCACATTCATTCACTTAACAATGAACTTGGGTCAAGGA 329  
QY 62 CGGCGTAGAGCTTACCTGTGTAGTAGTGAAGGCGGTGACAGTGTCTACATCTTTCA 121  
DB 330 CGGATGAGACTTACTGTGTATGAGGTGAGCGCATGCAATGACACCTGGGTCTG 389  
QY 122 CTGGAC 127  
DB 390 CTGAAC 395

## RESULT 15

US-10-460-923-1  
Sequence 1, Application US/10460923  
Publication No. US2004009951A1  
GENERAL INFORMATION:  
APPLICANT: MALIM, Michael H.  
APPLICANT: SHEEHY, Ann M.  
APPLICANT: HARRIS, Reuben S.  
APPLICANT: BISHOP, Kate N.  
APPLICANT: NEUBERGER, Michael S.  
APPLICANT: GADDIS, Nathan C.  
APPLICANT: SIMON, James H.M.  
TITLE OF INVENTION: DNA Deamination Mediates Innate Immunity to Retroviral Infection  
FILE REFERENCE: 22253-74380  
CURRENT APPLICATION NUMBER: US/10/460,923  
CURRENT FILING DATE: 2003-06-13  
PRIOR APPLICATION NUMBER: US 60/388,513  
PRIOR FILING DATE: 2002-06-13  
PRIOR APPLICATION NUMBER: US 60/472,952  
PRIOR FILING DATE: 2003-05-23  
NUMBER OF SEQ ID NOS: 12  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 1  
LENGTH: 1155  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-460-923-1

Query Match 27.8%; Score 41.2; DB 15; Length 1155;  
Best Local Similarity 57.9%; Pred. No. 5.8e-05;  
Matches 73; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 2 CTTTGATGAACCGAGAGAGTTCTTTACCAATTCAAAAATGTCGCTGCGCTAAGGGT 61  
DB 583 CACTCGATGATCCACCCACATTCATTCACTTAACAATGAACTTGGGTCAAGGA 642  
QY 62 CGGCGTAGAGCTTACCTGTGTAGTAGTGAAGGCGGTGACAGTGTCTACATCTTTCA 121  
DB 643 CGGATGAGACTTACTGTGTATGAGGTGAGCGCATGCAATGACACCTGGGTCTG 702  
QY 122 CTGGAC 127  
DB 703 CTGAAC 708

Search completed: March 13, 2004, 05:30:56  
Job time : 53.9361 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 12, 2004, 13:19:34 ; Search time 471.269 Seconds  
(without alignments)  
9378.080 Million cell updates/sec

Title: US-09-966-880A-12

Perfect score: 148

Sequence: 1 ccccttgatcgaccgagga.....ttggttcttcgcataag 148

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 segs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_estnu:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gse\_hum:\*  
18: em\_gse\_inv:\*  
19: em\_gse\_pln:\*  
20: em\_gse\_vrt:\*  
21: em\_gse\_fun:\*  
22: em\_gse\_mam:\*  
23: em\_gse\_mus:\*  
24: em\_gse\_pro:\*  
25: em\_gse\_rtd:\*  
26: em\_gse\_pig:\*  
27: em\_gse\_vrl:\*  
28: gb\_gse1:\*  
29: gb\_gse2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	148	100.0	541	10	BF238155 601811880
2	148	100.0	743	12	BG686133 602638412
3	148	100.0	820	12	BG757089 602715124
4	148	100.0	872	12	BG758510 602712721

5	148	100.0	889	12	BG686876 602650861
6	148	100.0	953	13	B0065440 AGENCOURT
7	148	100.0	1052	13	B0055935 AGENCOURT
8	148	100.0	1201	9	AL559877
9	148	100.0	1201	13	BX402063
10	146.4	98.9	535	14	CD707143
11	137	92.6	693	12	BG757392
12	137	92.6	942	10	BF975166
13	136	91.9	853	13	BX464579
14	117.6	79.5	701	28	BZ247020
15	93.6	63.2	623	9	AJ450295
16	93.6	63.2	650	9	AJ449745
17	93.6	63.2	688	9	AJ450317
18	93.6	63.2	696	9	AJ453647
19	93.6	63.2	729	9	AJ450296
20	93.6	63.2	757	9	AJ446140
21	54.2	36.6	852	29	CNS02108
22	45.8	30.9	358	13	BY012221
23	44.8	30.3	379	10	BB844013
24	44.8	30.3	771	10	BF141725
25	44.8	30.3	789	10	BF147755
26	44.8	30.3	925	10	BF138310
27	44.8	30.3	966	10	BF139214
28	43.2	29.2	359	13	BY100690
29	43.2	29.2	365	13	BY056132
30	43.2	29.2	605	10	BB651860
31	43.2	29.2	605	14	CA574608
32	43.2	29.2	646	13	BY738045
33	43.2	29.2	658	13	BY728644
34	43.2	29.2	663	10	BB638923
35	43.2	29.2	687	10	BB652985
36	43.2	29.2	772	12	BM935138
37	43.2	29.2	955	13	BQ934193
38	43.2	29.2	1202	11	AK049256
39	43.2	29.2	2226	11	AK049928
40	42.8	28.9	511	29	CG656381
41	42.8	28.9	1025	12	BM460629
42	42.4	28.6	358	10	BB841045
43	42.4	28.6	549	29	CG575900
44	42.2	28.5	522	12	BG144705
45	41.6	28.1	166	29	CG691593

## ALIGNMENTS

RESULT 1  
LOCUS BF238155 541 bp mRNA linear EST 14-NOV-2000  
DEFINITION 601811880F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4054915 5',  
mRNA sequence.  
ACCESSION BF238155  
VERSION BF238155.1 GI:11152074  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (Base 1 to 541)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.lnl.gov>  
Plate: LNCM895 row: p column: 20  
High quality sequence stop: 541.



FEATURES  
source

Location/Qualifiers  
1..541  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4054915"  
/issue\_type="primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_id="NIH\_MGC\_48"  
/note="Organ: B-cells; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC library."

## ORIGIN

Query Match 100.0%; Score 148; DB 10; Length 541;  
Best Local Similarity 100.0%; Pred. No. 9.7e-42;  
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTCTGATGACCGGAGAAAGTTCTTTACCAATTCAAAATGTCGGCTGAAGG 60  
DB 86 CCTCTGATGACCGGAGAAAGTTCTTTACCAATTCAAAATGTCGGCTGAAGG 145  
QY 61 TCGGCGTGAACCTACCTGCTACGTAAGAGAGCGTACAGTCAATCCTTTTC 120  
DB 146 TCGGCGTGAACCTACCTGCTACGTAAGAGAGCGTACAGTCAATCCTTTTC 205  
QY 121 ACTGACTTTGGTTATCTTCGCAATAG 148  
DB 206 ACTGACTTTGGTTATCTTCGCAATAG 233

RESULT 2  
BG686133 743 bp mRNA linear EST 01-MAY-2001  
LOCUS 60263842F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4766234 5',  
DEFINITION mRNA sequence.  
ACCESSION BG686133  
VERSION BG686133.1 GI:13917530  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 743)  
NIH-MGC http://mgc.nci.nih.gov/  
REFERENCE National Institutes of Health, Mammalian Gene Collection (MGC)  
AUTHORS Unpublished (1999)  
TITLE Contact: Robert Strausberg, Ph.D.  
JOURNAL Email: cgabs-remail.nih.gov  
COMMENT Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: LICM1626 row: 9 column: 03  
High quality sequence stop: 740.

## FEATURES

## source

Location/Qualifiers  
1..743  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4766234"  
/issue\_type="primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_id="NIH\_MGC\_48"

## ORIGIN

Query Match 100.0%; Score 148; DB 12; Length 743;  
Best Local Similarity 100.0%; Pred. No. 1.1e-41;  
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTCTGATGACCGGAGAAAGTTCTTTACCAATTCAAAATGTCGGCTGAAGG 60  
DB 86 CCTCTGATGACCGGAGAAAGTTCTTTACCAATTCAAAATGTCGGCTGAAGG 145  
QY 61 TCGGCGTGAACCTACCTGCTACGTAAGAGAGCGTACAGTCAATCCTTTTC 120  
DB 146 TCGGCGTGAACCTACCTGCTACGTAAGAGAGCGTACAGTCAATCCTTTTC 205  
QY 121 ACTGACTTTGGTTATCTTCGCAATAG 148  
DB 206 ACTGACTTTGGTTATCTTCGCAATAG 233

RESULT 3  
BG757089 820 bp mRNA linear EST 15-MAY-2001  
LOCUS 602715124F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4855517 5',  
DEFINITION mRNA sequence.  
ACCESSION BG757089  
VERSION BG757089.1 GI:14067742  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 820)  
NIH-MGC http://mgc.nci.nih.gov/  
REFERENCE National Institutes of Health, Mammalian Gene Collection (MGC)  
AUTHORS Unpublished (1999)  
TITLE Contact: Robert Strausberg, Ph.D.  
JOURNAL Email: cgabs-remail.nih.gov  
COMMENT Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: LICM1704 row: 0 column: 06  
High quality sequence stop: 675.

FEATURES  
source

Location/Qualifiers  
1..820  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4855517"  
/issue\_type="primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_id="NIH\_MGC\_48"  
/note="Organ: B-cells; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC library."



## ORIGIN

Query Match 100.0%; Score 148; DB 12; Length 820;  
Best Local Similarity 100.0%; Pred. No. 1.2e-41;  
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTCTGATGACCGGAGGAAGTTCTTTACCAATTCAAAATGTCGGCTGAGG 60  
DB CCTCTGATGACCGGAGGAAGTTCTTTACCAATTCAAAATGTCGGCTGAGG 152  
QY 61 TCGGCGTAGACCTACCTGCTGCTAGTAGAAGAGCGTGACAGTCTACATCTTTTC 120  
DB 153 TCGGCGTAGACCTACCTGCTGCTAGTAGAAGAGCGTGACAGTCTACATCTTTTC 212  
QY 121 ACTGACCTTGCTATCTTCCCAATAG 148  
DB 213 ACTGACCTTGCTATCTTCCCAATAG 240

## RESULT 4

BG758510

LOCUS 872 bp mRNA linear EST 15-MAY-2001

DEFINITION 60271272.F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4853069 5',

mRNA sequence.

ACCESSION BG758510

VERSION BG758510.1 GI:14069163

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 872)

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs-remail.nih.gov

Tissue Procurement: Louis M. Staudt, M.D., Ph.D.

CDNA Library Preparation: Ling Hong/Rubin Laboratory

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LCM1698 row: i column: 06

High quality sequence stop: 836.

Location/Qualifiers

1..872

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:4853069"

/tissue\_type="primary B-cells from tonsils (cell line)"

/lab\_host="DH10B (phage-resistant)"

/clone\_id="NIH\_MGC\_48"

/note="Organ: B-cells; Vector: pOTB7; Site 1: XhoI;

Site 2: EcoRI; cDNA made by oligo-dT priming.

Directionally cloned into EcoRI/XhoI sites using the

following 5' adaptor: GGCAAGAG(G). Size-selected &gt;500bp

for average insert size 1.8kb. Library constructed by Ling

Hong in the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies).

Note: this is a NIH\_MGC Library."

Query Match 100.0%; Score 148; DB 12; Length 872;

Best Local Similarity 100.0%; Pred. No. 1.2e-41;

Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTCTGATGACCGGAGGAAGTTCTTTACCAATTCAAAATGTCGGCTGAGG 60  
DB 74 CCTCTGATGACCGGAGGAAGTTCTTTACCAATTCAAAATGTCGGCTGAGG 133

QY 61 TCGGCGTAGACCTACCTGCTGCTAGTAGAAGAGCGTGACAGTCTACATCTTTTC 120  
DB 134 TCGGCGTAGACCTACCTGCTGCTAGTAGAAGAGCGTGACAGTCTACATCTTTTC 193

QY 121 ACTGACCTTGCTATCTTCCCAATAG 148  
DB 194 ACTGACCTTGCTATCTTCCCAATAG 221

## RESULT 5

BG686876

LOCUS 889 bp mRNA linear EST 01-MAY-2001

DEFINITION 60265086.F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4763247 5',

mRNA sequence.

ACCESSION BG686876

VERSION BG686876.1 GI:13918273

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 889)

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs-remail.nih.gov

Tissue Procurement: Louis M. Staudt, M.D., Ph.D.

CDNA Library Preparation: Ling Hong/Rubin Laboratory

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LCM1618 row: j column: 16

High quality sequence start: 6

High quality sequence stop: 727.

Location/Qualifiers

1..889

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:4763247"

/tissue\_type="primary B-cells from tonsils (cell line)"

/lab\_host="DH10B (phage-resistant)"

/clone\_id="NIH\_MGC\_48"

/note="Organ: B-cells; Vector: pOTB7; Site 1: XhoI;

Site 2: EcoRI; cDNA made by oligo-dT priming.

Directionally cloned into EcoRI/XhoI sites using the

following 5' adaptor: GGCAAGAG(G). Size-selected &gt;500bp

for average insert size 1.8kb. Library constructed by Ling

Hong in the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies).

Note: this is a NIH\_MGC Library."

Query Match 100.0%; Score 148; DB 12; Length 889;

Best Local Similarity 100.0%; Pred. No. 1.2e-41;

Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTCTGATGACCGGAGGAAGTTCTTTACCAATTCAAAATGTCGGCTGAGG 60  
DB 46 CCTCTGATGACCGGAGGAAGTTCTTTACCAATTCAAAATGTCGGCTGAGG 105  
QY 61 TCGGCGTAGACCTACCTGCTGCTAGTAGAAGAGCGTGACAGTCTACATCTTTTC 120  
DB 106 TCGGCGTAGACCTACCTGCTGCTAGTAGAAGAGCGTGACAGTCTACATCTTTTC 165  
QY 121 ACTGACCTTGCTATCTTCCCAATAG 148  
DB 166 ACTGACCTTGCTATCTTCCCAATAG 193



RESULT 6  
LOCUS BQ065440  
DEFINITION AGENCOURT\_6855061 NIH\_MGC\_99 Homo sapiens cDNA clone IMAGE:5929977  
5', mRNA sequence.  
ACCESSION BQ065440  
VERSION BQ065440.1 GI:19894486  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 953)  
NIH-MGC <http://mgs.nci.nih.gov/>  
National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: Lou Staudt  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
DNA distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.lnl.gov>  
Plate: LNCM2108 row: p column: 10  
High quality sequence stop: 634.  
Location/Qualifiers  
1. 953  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5929977"  
/issue\_type="lymphoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_99"  
/note="Organ: lymph; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCAGCAG(G). Size-selected >500bp for average insert size  
1.8kb. Library constructed by Ling Hong in the laboratory  
of Gerald M. Rubin (University of California, Berkeley)  
using ZAP-cDNA synthesis kit (Stratagene) and Superscript  
II RT (Life Technologies). Note: this is a NIH\_MGC  
Library."

ORIGIN  
Query Match 100.0%; Score 148; DB 13; Length 953;  
Best Local Similarity 100.0%; Pred. No. 1.3e-41;  
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CCTCTGATGAACCGAGAGAGTTTCTTACCAATCAAAATGTCGCTGGGCTAAGG 60  
DB 72 CCTCTGATGAACCGAGAGAGTTTCTTACCAATCAAAATGTCGCTGGGCTAAGG 131  
QY 61 TCGGCGTGAACCTACCTGTGCTACGTAGTGAAGAGCGCTGACAGTCTACATCCTTTTC 120  
DB 132 TCGGCGTGAACCTACCTGTGCTACGTAGTGAAGAGCGCTGACAGTCTACATCCTTTTC 191  
QY 121 ACTGACTTTGGTTATCTTTCGCAATAAG 148  
DB 192 ACTGACTTTGGTTATCTTTCGCAATAAG 219

RESULT 7  
LOCUS BQ055935  
DEFINITION AGENCOURT\_6796291 NIH\_MGC\_99 Homo sapiens cDNA clone IMAGE:5808181  
5', mRNA sequence.  
ACCESSION BQ055935  
VERSION BQ055935.1 GI:19815262

KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1052)  
NIH-MGC <http://mgs.nci.nih.gov/>  
National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: Lou Staudt  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
DNA distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.lnl.gov>  
Plate: LNCM2051 row: m column: 14  
High quality sequence stop: 665.  
Location/Qualifiers  
1. 1052  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5808181"  
/issue\_type="lymphoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_99"  
/note="Organ: lymph; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCAGCAG(G). Size-selected >500bp for average insert size  
1.8kb. Library constructed by Ling Hong in the laboratory  
of Gerald M. Rubin (University of California, Berkeley)  
using ZAP-cDNA synthesis kit (Stratagene) and Superscript  
II RT (Life Technologies). Note: this is a NIH\_MGC  
Library."

ORIGIN  
Query Match 100.0%; Score 148; DB 13; Length 1052;  
Best Local Similarity 100.0%; Pred. No. 1.3e-41;  
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CCTTTATGAACCGAGAGAGTTTCTTACCAATCAAAATGTCGCTGGGCTAAGG 60  
DB 72 CCTTTATGAACCGAGAGAGTTTCTTACCAATCAAAATGTCGCTGGGCTAAGG 131  
QY 61 TCGGCGTGAACCTACCTGTGCTACGTAGTGAAGAGCGCTGACAGTCTACATCCTTTTC 120  
DB 132 TCGGCGTGAACCTACCTGTGCTACGTAGTGAAGAGCGCTGACAGTCTACATCCTTTTC 191  
QY 121 ACTGACTTTGGTTATCTTTCGCAATAAG 148  
DB 192 ACTGACTTTGGTTATCTTTCGCAATAAG 219

RESULT 8  
LOCUS AL559877  
DEFINITION AL559877 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA  
clone CS0D00031B14 5-PRIME, mRNA sequence.  
ACCESSION AL559877  
VERSION AL559877.2 GI:31284008  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1201)  
Li, W. B., Gruber, C., Jessee, J., and Polyes, D.  
Full-length cDNA libraries and normalization  
unpublished (2001)







Db 129 CCTCTGATGAAACCGAGGAAGTTCTTACCAATTCAAAATGTCGGCTGGGCTAAGG 188

QY 61 TCGGCGTGAAGCCTACCTGCTACGTAGTAAGAGCGCTGACAGTCTACATCTTTTC 120

Db 189 TCGGCGTGAAGCCTACCTGCTACGTAGTAAGAGCGCTGACAGTCTACATCTTTTC 248

QY 121 ACTGACCTTGGTTATCTTCCCAATTAAG 148

Db 249 ACTGACCTTGGTTATCTTCCCAATTAAG 276

RESULT 11  
Bg757392 693 bp mRNA linear EST 15-MAY-2001  
LOCUS 602711022P1 NIH\_MGC\_48 Homo sapiens CDNA clone IMAGE:4851580 5',  
DEFINITION mRNA sequence.  
ACCESSION Bg757392  
VERSION Bg757392  
KEYWORDS EST. GI:14068045  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE NIH-MGC http://mgc.nci.nih.gov/  
AUTHORS 1 (bases 1 to 693)  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: LNCM1694 row: k column: 05  
High quality sequence stop: 693.  
Location/Qualifiers  
1..693  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4851580"  
/tissue\_type="primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_48"  
/note="Organ: B-cells; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

FEATURES  
source

Query Match 92.6%; Score 137; DB 12; Length 693;  
Best Local Similarity 99.3%; Pred. No. 9.5e-38;  
Matches 148; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 CCTCTGATGAAACCGAGGAAGTTCTTACCAATTCAAAATGTCGGCTGGGCTAAGG 60

Db 89 CCTCTGATGAAACCGAGGAAGTTCTTACCAATTCAAAATGTCGGCTGGGCTAAGG 148

QY 61 TCGGCGTGAAGCCTACCTGCTACGTAGTAAGAGCGCTGACAGTCTACATCTTTTC 119

Db 149 TCGGCGTGAAGCCTACCTGCTACGTAGTAAGAGCGCTGACAGTCTACATCTTTTC 208

QY 120 CACTGACCTTGGTTATCTTCCCAATTAAG 148

Db 209 CACTGACCTTGGTTATCTTCCCAATTAAG 237

RESULT 12  
BF975166 942 bp mRNA linear EST 22-JAN-2001  
LOCUS 602244657P1 NIH\_MGC\_48 Homo sapiens CDNA clone IMAGE:4335639 5',  
DEFINITION mRNA sequence.  
ACCESSION BF975166  
VERSION BF975166  
KEYWORDS EST. GI:12342381  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE NIH-MGC http://mgc.nci.nih.gov/  
AUTHORS 1 (bases 1 to 942)  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: LNCM1207 row: a column: 16  
High quality sequence stop: 707.  
Location/Qualifiers  
1..942  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4335639"  
/tissue\_type="primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_48"  
/note="Organ: B-cells; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

FEATURES  
source

Query Match 92.6%; Score 137; DB 10; Length 942;  
Best Local Similarity 99.3%; Pred. No. 1.1e-37;  
Matches 148; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 CCTCTGATGAAACCGAGGAAGTTCTTACCAATTCAAAATGTCGGCTGGGCTAAGG 59

Db 86 CCTCTGATGAAACCGAGGAAGTTCTTACCAATTCAAAATGTCGGCTGGGCTAAGG 145

QY 60 GTGCGCGTGAAGCCTACCTGCTACGTAGTAAGAGCGCTGACAGTCTACATCTTTTC 119

Db 146 GTGCGCGTGAAGCCTACCTGCTACGTAGTAAGAGCGCTGACAGTCTACATCTTTTC 205

QY 120 CACTGACCTTGGTTATCTTCCCAATTAAG 148

Db 206 CACTGACCTTGGTTATCTTCCCAATTAAG 234

RESULT 13  
BX464579 853 bp mRNA linear EST 22-MAY-2003  
LOCUS BX464579 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens CDNA clone CS0D0003YB14 5-PRIME, mRNA sequence.  
ACCESSION BX464579



VERSION BX464579.1 GI:31031641  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 853)  
AUTHORS Li, W.B., Gruber, C., Jesse, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: [secref@genoscope.cns.fr](mailto:secref@genoscope.cns.fr), Web: [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 6672.r for  
more information about this cluster, see  
<http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS1DG001ZF10CP1&cluster=6672.r>. Contact :  
Feng Ling Email : [fliang@life.com](mailto:fliang@life.com) URL :  
<http://fulllength.invitrogen.com/> Invitrogen Corporation 1600  
Paraday Avenue Genoscope sequence ID : CS1DG001ZF10CP1.  
Location/Qualifiers  
source  
1. 853  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0G003YB14"  
/tissue\_type="B CELLS (RAMOS CELL LINE)"  
/cell\_line="RAMOS CELL LINE"  
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed  
with a NotI-oligo(dT) primer. Five prime end enriched,  
double-strand cDNA was digested with Not I and cloned into  
the Not I and EcoRV sites of the pCMVSPORT 6 vector.  
Library was not normalized."

ORIGIN  
Query Match 91.9%; Score 136; DB 13; Length 853;  
Best Local Similarity 99.3%; Pred. No. 2.4e-37;  
Matches 147; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
QY 1 CCTCTGATGAGACCGAGAGATTCTTTACCAATTCAAAATGTCCTGGGCTAAGG 60  
DB 95 CCTCTGATGAGACCGAGAGATTCTTTACCAATTCAAAATGTCCTGGGCTAAGG 153  
QY 61 TCGGCGTGAAGACCTACCTGCTAGTAGAGAGAGCGTGAAGTCTACATCCTTTTC 120  
DB 154 TCGGCGTGAAGACCTACCTGCTAGTAGAGAGAGCGTGAAGTCTACATCCTTTTC 213  
QY 121 ACTGACTTTGTTATCTTCGCAATAG 148  
DB 214 ACTGACTTTGTTATCTTCGCAATAG 241  
RESULT 14  
BZ247020/c 701 bp DNA linear GSS 12-OCT-2002  
LOCUS CH230-374G16.T3 CHORI-230 Segment 2 Rattus norvegicus genomic clone  
DEFINITION BZ247020  
ACCESSION BZ247020  
VERSION BZ247020.1 GI:23907284  
KEYWORDS GSS.  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE 1 (bases 1 to 701)  
AUTHORS Zhao, S., Shetty, J., Shatsman, S., Teagay, G., Geer, K.,  
Rivartsbeyn, A., Geirgeorgis, E., Overton, L., Russell, D., Chen, D.,  
Riggs, F., de Jong, P. and Fraser, C.M.  
TITLE Rat BAC End Sequences from Library CHORI-230 MboI segment

JOURNAL Unpublished (1999)  
COMMENT Other GSSs: CH230-374G16.TV  
Contact: Shaying Zhao  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: [szhao@tigr.org](mailto:szhao@tigr.org)  
Clones are derived from the rat BAC library CHORI-230  
(<http://www.chori.org/bacpac/rat230.htm>). For BAC library  
availability, please contact Pieter de Jong ([pdjong@mail.cho.org](mailto:pdjong@mail.cho.org)).  
Clones may be purchased from BACPAC Resources  
(<http://www.chori.org/bacpac/or> ering information.htm). BAC end  
page: [http://www.tigr.org/tftp/bac\\_ends/rat/bac\\_end\\_intro.html](http://www.tigr.org/tftp/bac_ends/rat/bac_end_intro.html)  
Plate: 374 row G column: 16  
Seq primer: Sp6  
Class: BAC ends.  
Location/Qualifiers  
source  
1. 701  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/strain="BN/SSNHsd/MCW"  
/db\_xref="taxon:10116"  
/clone="CH230-374G16"  
/sex="Female"  
/cell\_type="Brain"  
/clone\_id="CHORI-230 Segment 2"  
/note="Vector: PTARBAC1.3; Site 1: MboI; site 2: MboI;  
CHORI-230 Rat (BN/SSNHsd/MCW) BAC library produced by  
Pieter de Jong"

ORIGIN  
Query Match 79.5%; Score 117.6; DB 28; Length 701;  
Best Local Similarity 87.2%; Pred. No. 8.5e-31;  
Matches 129; Conservative 0; Mismatches 19; Indels 0; Gaps 0;  
QY 1 CCTTTGATGACCGAGAGATTCTTTACCAATTCAAAATGTCCTGGGCTAAGG 60  
DB 602 CCTTTGATGAGAGAGAGAGATTCTTTACCACTTCAAAATGTCCTGGGCTAAGG 543  
QY 61 TCGGCGTGAAGACCTACCTGCTAGTAGAGAGAGCGTGAAGTCTACATCCTTTTC 120  
DB 542 TCGGCGTGAAGACCTACCTGCTAGTAGAGAGAGCGTGAAGTCTACATCCTTTTC 483  
QY 121 ACTGACTTTGTTATCTTCGCAATAG 148  
DB 482 ACTGACTTTGCGCACTTGGCAACAG 455  
RESULT 15  
A450295 623 bp mRNA linear EST 19-APR-2002  
LOCUS A450295  
DEFINITION A450295 riken1 Gallus gallus cDNA clone 24m21r1, mRNA sequence.  
ACCESSION A450295  
VERSION A450295.1 GI:20217516  
KEYWORDS EST.  
SOURCE Gallus gallus (chicken)  
ORGANISM Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
REFERENCE 1 (bases 1 to 623)  
AUTHORS Buerstedde, J.M.  
TITLE Gallus gallus bursal lymphocyte EST  
JOURNAL Unpublished (2002)  
COMMENT Contact: Buerstedde JM  
Cellular Immunology  
Heinrich-Pette-Institute  
Martiniestr. 52, 20251 Hamburg, Germany  
Email: [URL: http://genetics.hpi.uni-hamburg.de/dt40est.html](mailto:URL: http://genetics.hpi.uni-hamburg.de/dt40est.html).  
Location/Qualifiers  
source  
1. 623



/organism="Gallus gallus"  
/mol\_type="mRNA"  
/db\_xref="taxon:9031"  
/clone="24m21x1"  
/cell\_type="bursal lymphocyte"  
/dev\_stage="2-3 weeks old"  
/clone\_id="r1ken1"  
/note="CB inbred strain"

## ORIGIN

Query Match 63.2%; Score 93.6; DB 9; Length 623;  
Best Local Similarity 77.0%; Pred. No. 3.2e-22;  
Matches 114; Conservative 0; Mismatches 34; Indels 0; Gaps 0;  
QY 1 CCTCTGATGAACCGAGGAAGTTCTTACCAATCAAAAATGCCGTGGCTAAGG 60  
DB 165 CCTCTGATGAAGAGAGAGCTCTTCTCTACAATTTCAGAACTGCCGTGGCCAAAG 224  
QY 61 TCGGCGTGAAGCCTTCTGTCTACGTAGTGAAGGCGGTGAACGTCTACATCCTTTTC 120  
DB 225 CCGTCGTGAACCTACCTCTGTATGTGTGAAGCGCGGTGAACGTCTACATCATGCTC 284  
QY 121 ACTGCACTTGTATTATCTTGCATTAAG 148  
DB 285 CCTGCACTTGTATACCTGCGTAACAAG 312

Search completed: March 13, 2004, 00:39:02  
Job time : 474.394 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

# OM nucleic - nucleic search, using sw model

Run on: March 12, 2004, 12:27:59 ; Search time 853.181 Seconds

(without alignments)  
13767.261 Million cell updates/sec

Title: US-09-966-880A-13

Perfect score: 271

Sequence: 1 aaagcgtcgcacgttggaatt.....agccatcatcaccctcaag 271

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenBank:\*

1: gb\_ba:\*

2: gb\_hg:\*

3: gb\_in:\*

4: gb\_om:\*

5: gb\_ov:\*

6: gb\_pat:\*

7: gb\_ph:\*

8: gb\_pl:\*

9: gb\_pr:\*

10: gb\_ro:\*

11: gb\_sts:\*

12: gb\_sy:\*

13: gb\_un:\*

14: gb\_vl:\*

15: em\_ba:\*

16: em\_fun:\*

17: em\_hum:\*

18: em\_in:\*

19: em\_mu:\*

20: em\_om:\*

21: em\_or:\*

22: em\_ov:\*

23: em\_pat:\*

24: em\_ph:\*

25: em\_pl:\*

26: em\_ro:\*

27: em\_sts:\*

28: em\_un:\*

29: em\_vl:\*

30: em\_hg\_hum:\*

31: em\_hg\_inv:\*

32: em\_hg\_other:\*

33: em\_hg\_mus:\*

34: em\_hg\_pln:\*

35: em\_hg\_rtd:\*

36: em\_hg\_man:\*

37: em\_hg\_vtl:\*

38: em\_sy:\*

39: em\_hgo\_hum:\*

40: em\_hgo\_mus:\*

41: em\_hgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	271	100.0	271	6 BD016838	BD016838 Novel cyt
2	271	100.0	547	12 AF529856	AF529856 Cricetulu
3	271	100.0	591	12 AF529855	AF529855 Cricetulu
4	271	100.0	595	12 AF529834	AF529834 Mus muscu
5	271	100.0	596	9 AF529815	AF529815 Homo sapi
6	271	100.0	596	9 AF529822	AF529822 Homo sapi
7	271	100.0	596	9 AF529823	AF529823 Homo sapi
8	271	100.0	596	9 AF529825	AF529825 Homo sapi
9	271	100.0	596	12 AF529829	AF529829 Mus muscu
10	271	100.0	596	12 AF529831	AF529831 Mus muscu
11	271	100.0	596	12 AF529832	AF529832 Mus muscu
12	271	100.0	596	12 AF529833	AF529833 Mus muscu
13	271	100.0	596	12 AF529835	AF529835 Mus muscu
14	271	100.0	596	12 AF529836	AF529836 Mus muscu
15	271	100.0	597	9 BT007402	BT007402 Homo sapi
16	271	100.0	597	12 AF529842	AF529842 Cricetulu
17	271	100.0	597	12 AF529843	AF529843 Cricetulu
18	271	100.0	597	12 AF529846	AF529846 Cricetulu
19	271	100.0	597	12 AF529847	AF529847 Cricetulu
20	271	100.0	597	12 AF529851	AF529851 Cricetulu
21	271	100.0	597	12 AF529853	AF529853 Cricetulu
22	271	100.0	597	12 BT008226	BT008226 Synthetic
23	271	100.0	1828	9 BC006296	BC006296 Homo sapi
24	271	100.0	2791	9 AB040431	AB040431 Homo sapi
25	271	100.0	2818	6 BD016833	BD016833 Novel cyt
26	271	100.0	6564	6 BD016835	BD016835 Novel cyt
27	271	100.0	11204	6 BD016860	BD016860 Novel cyt
28	271	100.0	11204	9 AB040430	AB040430 Homo sapi
29	271	100.0	71132	9 AC092184	AC092184 Homo sapi
30	269.4	99.4	577	12 AF529840	AF529840 Mus muscu
31	269.4	99.4	596	9 AF529816	AF529816 Homo sapi
32	269.4	99.4	596	9 AF529817	AF529817 Homo sapi
33	269.4	99.4	596	9 AF529818	AF529818 Homo sapi
34	269.4	99.4	596	9 AF529819	AF529819 Homo sapi
35	269.4	99.4	596	9 AF529820	AF529820 Homo sapi
36	269.4	99.4	596	9 AF529821	AF529821 Homo sapi
37	269.4	99.4	596	9 AF529824	AF529824 Homo sapi
38	269.4	99.4	596	9 AF529826	AF529826 Homo sapi
39	269.4	99.4	596	9 AF529827	AF529827 Homo sapi
40	269.4	99.4	596	12 AF529830	AF529830 Mus muscu
41	269.4	99.4	596	12 AF529837	AF529837 Mus muscu
42	269.4	99.4	596	12 AF529839	AF529839 Mus muscu
43	269.4	99.4	597	12 AF529828	AF529828 Mus muscu
44	269.4	99.4	597	12 AF529841	AF529841 Cricetulu
45	269.4	99.4	597	12 AF529844	AF529844 Cricetulu

## ALIGNMENTS

RESULT 1

BD016838

LOCUS: BD016838

DEFINITION: Novel cytidine deaminase.

ACCESSION: BD016838

VERSION: BD016838.1 GI:22558014

KEYWORDS: UP 2001245669-A/11.

SOURCE: Homo sapiens (human)

ORGANISM: Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 271)

AUTHORS: Honjo, T. and Muramatsu, M.

TITLE: Novel cytidine deaminase

JOURNAL: Patent: JP 2001245669-A 11 11-SEP-2001;



COMMENT JAPAN TOBACCO INC, TASUKU HONJO  
OS Homo sapiens (human)  
PN JP 2001245669-A/11  
PD 11-SEP-2001  
PR 28-MAR-2000 JP 2000092981  
PI TASUKU HONJO, MASAMICHI MURAMATSU  
PC C12N15/09, A61K39/395, A61P1/00, A61P11/06, A61P13/12,  
PC A61P17/00,  
PC A61P27/02, A61P27/16, A61P37/02, A61P37/08, C07K16/18, C12N1/19, PC  
C12N1/21,  
PC C12N5/10, C12N9/78, C12P21/02, C12P21/08, C12N1/21, C12R1/19, PC  
(C12N5/10, C12R1/91), C12N15/00, C12N5/00, C12N5/00, C12R1/91) CC  
FH Key Location/Qualifiers.  
1. 271  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

ORIGIN  
Query Match 100.0%; Score 271; DB 6; Length 271;  
Best Local Similarity 100.0%; Pred. No. 6, 4e-50;  
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACGGCTGCCACGTGGAAATGCTCTTCCCTCCGCTACATCTCGGACTGGAGCTAGACCT 60  
Db 1 AACGGCTGCCACGTGGAAATGCTCTTCCCTCCGCTACATCTCGGACTGGAGCTAGACCT 60  
QY 61 GGCGCTGCTACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120  
Db 61 GGCGCTGCTACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120  
QY 121 CATGTGGCCGCACTTTCTGCGAGAGGAAACCCCACTCACTGAGATCTTCAACCGGCGCC 180  
Db 121 CATGTGGCCGCACTTTCTGCGAGAGGAAACCCCACTCACTGAGATCTTCAACCGGCGCC 180  
QY 181 CTCTACTCTGTGTAGAGACCGGAGGCTGAGCCCGAGGGGCTGGCGGCTGCAACCGCGCC 240  
Db 181 CTCTACTCTGTGTAGAGACCGGAGGCTGAGCCCGAGGGGCTGGCGGCTGCAACCGCGCC 240  
QY 241 GGGGTGCAATAGCCATCATGACCTTCAAG 271  
Db 241 GGGGTGCAATAGCCATCATGACCTTCAAG 271

RESULT 2  
AF529856 547 bp mRNA linear SYN 21-AUG-2003  
LOCUS Cricetus griseus transgenic clone 16 Homo sapiens AID (AID) mRNA,  
DEFINITION complete cds.  
ACCESSION AF529856  
VERSION AF529856.1 GI:22297297  
KEYWORDS  
SOURCE Cricetus griseus (Chinese hamster)  
ORGANISM Cricetus griseus (Chinese hamster)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
Cricetulus.  
1 (bases 1 to 547)  
Martin, A. and Scharff, M.D.  
Somatic hypermutation of the AID transgene in B and non-B cells  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12304-12308 (2002)  
MEDLINE 22222991  
PUBMED 12202747  
REFERENCE 2 (bases 1 to 547)  
AUTHORS Martin, A. and Scharff, M.D.  
TITLE Direct Submission  
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of  
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA  
FEATURES  
1. 547  
Location/Qualifiers  
/organism="Cricetus griseus"  
/mol\_type="mRNA"

SOURCE /db\_xref="taxon:10029"  
/clone="16"  
/cell\_line="CHO"  
/transgenic  
1. 547  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
1. 547  
/gene="AID"  
1. 547  
/gene="AID"  
/codon\_start=2  
/product="AID"  
/protein\_id="AA95443.1"  
/db\_xref="GI:22297300"  
/translation="VRKAGRETYCYVVKRSDATSPSIDPGLRNNGCHELLE  
LRIYSDMDLPGRCYRYTFTMSPTCDKAPVADFLRGNPISLRIFFARLYPCEDR  
KASPEGLRLHRAVOIAINTFYKWTVEVNERITFKAMEGLHENSVRLSRLR  
RIILFLEYVDLDRDAFRWGR"

ORIGIN  
Query Match 100.0%; Score 271; DB 12; Length 547;  
Best Local Similarity 100.0%; Pred. No. 5, 9e-50;  
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACGGCTGCCACGTGGAAATGCTCTTCCCTCCGCTACATCTCGGACTGGAGCTAGACCT 60  
Db 107 AACGGCTGCCACGTGGAAATGCTCTTCCCTCCGCTACATCTCGGACTGGAGCTAGACCT 166  
QY 61 GGCGCTGCTACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120  
Db 167 GGCGCTGCTACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 226  
QY 121 CATGTGGCCGCACTTTCTGCGAGAGGAAACCCCACTCACTGAGATCTTCAACCGGCGCC 180  
Db 121 CATGTGGCCGCACTTTCTGCGAGAGGAAACCCCACTCACTGAGATCTTCAACCGGCGCC 180  
QY 227 CATGTGGCCGCACTTTCTGCGAGAGGAAACCCCACTCACTGAGATCTTCAACCGGCGCC 286  
Db 227 CATGTGGCCGCACTTTCTGCGAGAGGAAACCCCACTCACTGAGATCTTCAACCGGCGCC 286  
QY 181 CTCTACTCTGTGTAGAGACCGGAGGCTGAGCCCGAGGGGCTGGCGGCTGCAACCGGCGCC 240  
Db 181 CTCTACTCTGTGTAGAGACCGGAGGCTGAGCCCGAGGGGCTGGCGGCTGCAACCGGCGCC 240  
QY 287 CTCTACTCTGTGTAGAGACCGGAGGCTGAGCCCGAGGGGCTGGCGGCTGCAACCGGCGCC 346  
Db 287 CTCTACTCTGTGTAGAGACCGGAGGCTGAGCCCGAGGGGCTGGCGGCTGCAACCGGCGCC 346  
QY 241 GGGGTGCAATAGCCATCATGACCTTCAAG 271  
Db 241 GGGGTGCAATAGCCATCATGACCTTCAAG 271  
QY 347 GGGGTGCAATAGCCATCATGACCTTCAAG 377  
Db 347 GGGGTGCAATAGCCATCATGACCTTCAAG 377

RESULT 3  
AF529855 591 bp mRNA linear SYN 21-AUG-2003  
LOCUS Cricetus griseus transgenic clone 15 Homo sapiens AID (AID) mRNA,  
DEFINITION complete cds.  
ACCESSION AF529855  
VERSION AF529855.1 GI:22297297  
KEYWORDS  
SOURCE Cricetus griseus (Chinese hamster)  
ORGANISM Cricetus griseus (Chinese hamster)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
Cricetulus.  
1 (bases 1 to 591)  
Martin, A. and Scharff, M.D.  
Somatic hypermutation of the AID transgene in B and non-B cells  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12304-12308 (2002)  
MEDLINE 22222991  
PUBMED 12202747  
REFERENCE 2 (bases 1 to 591)  
AUTHORS Martin, A. and Scharff, M.D.  
TITLE Direct Submission  
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of  
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA  
FEATURES  
1. 591  
Location/Qualifiers







gene  
1. .>596  
/gene="AID"  
CDS  
1. .>596  
/gene="AID"  
/note="Integrated into Burkitt's lymphoma cell line Ramos"  
/codon\_start=1  
/product="AID"  
/protein\_id="AAM95402.1"  
/db\_xref="GI:22297218"  
/translation="MDSLMMNRKFLYQKNVWAKGRRETYLCYVWRDASPSFL  
DFGLRNKNGCHEVLLFLRYISDMDDPGRCTRVTFWTFMSPTCDARHVADELGNP  
NLSLRIFTLARYFCEDRKAPBEGRLHRAQVQIALMTFKDYFCMNTVEVNERHRTFK  
AMEGLHENSVRLSRQLRRIILPLVEYDDLRDARFTLGL"

## ORIGIN

Query Match 100.0%; Score 271; DB 9; Length 596;  
Best Local Similarity 100.0%; Pred. No. 5,9e-50;  
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACGGCTGCCAGCTGGAATGCTCTTCCTCCGCTACATCTCGAGACTGGACCTAGACCT 60  
157 AACGGCTGCCAGCTGGAATGCTCTTCCTCCGCTACATCTCGAGACTGGACCTAGACCT 216  
QY 61 GGCCGCTGCTACCGCGCTCACTGCTTCACTCTGAGAGCCCTGCTACGACTGTGCCGA 120  
217 GGCCGCTGCTACCGCGCTCACTGCTTCACTCTGAGAGCCCTGCTACGACTGTGCCGA 276  
QY 121 CATGTGGCCGACTTTTCTGGAGAGAAACCCCAACTCAGTCTGAGATCTTCAACGGCGGC 180  
277 CATGTGGCCGACTTTTCTGGAGAGAAACCCCAACTCAGTCTGAGATCTTCAACGGCGGC 336  
QY 181 CTCTACTTCTGTGAGAGACCGCAAGGCTGAGCCCGAGGGCTGCGGGCTGACACGGCGCC 240  
337 CTCTACTTCTGTGAGAGACCGCAAGGCTGAGCCCGAGGGCTGCGGGCTGACACGGCGCC 396  
QY 241 GGGGTGCAAAATAGCATCATGACCTTCAAG 271  
397 GGGGTGCAAAATAGCATCATGACCTTCAAG 427  
DB

RESULT 6  
AF529822 596 bp mRNA linear PRI 17-SEP-2002  
LOCUS Homo sapiens clone Ramos 8 AID (AID) mRNA, partial cds.  
DEFINITION  
ACCESSION AF529822  
VERSION AF529822.1 GI:22297231  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Martin,A. and Scharff,M.D.  
TITLE Somatic hypermutation of the AID transgene in B and non-B cells  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12304-12308 (2002)  
REFERENCE  
AUTHORS Martin,A. and Scharff,M.D.  
TITLE Direct Submmission  
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of  
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA  
FEATURES  
Location/Qualifiers  
source  
1. .>596  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="Ramos 8"  
1. .>596  
/gene="AID"  
1. .>596  
/gene="AID"  
/note="Integrated into Burkitt's lymphoma cell line Ramos"  
/codon\_start=1

## ORIGIN

Query Match 100.0%; Score 271; DB 9; Length 596;  
Best Local Similarity 100.0%; Pred. No. 5,9e-50;  
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACGGCTGCCAGCTGGAATGCTCTTCCTCCGCTACATCTCGAGACTGGACCTAGACCT 60  
157 AACGGCTGCCAGCTGGAATGCTCTTCCTCCGCTACATCTCGAGACTGGACCTAGACCT 216  
QY 61 GGCCGCTGCTACCGCGCTCACTGCTTCACTCTGAGAGCCCTGCTACGACTGTGCCGA 120  
217 GGCCGCTGCTACCGCGCTCACTGCTTCACTCTGAGAGCCCTGCTACGACTGTGCCGA 276  
QY 121 CATGTGGCCGACTTTTCTGGAGAGAAACCCCAACTCAGTCTGAGATCTTCAACGGCGGC 180  
277 CATGTGGCCGACTTTTCTGGAGAGAAACCCCAACTCAGTCTGAGATCTTCAACGGCGGC 336  
QY 181 CTCTACTTCTGTGAGAGACCGCAAGGCTGAGCCCGAGGGCTGCGGGCTGACACGGCGCC 240  
337 CTCTACTTCTGTGAGAGACCGCAAGGCTGAGCCCGAGGGCTGCGGGCTGACACGGCGCC 396  
QY 241 GGGGTGCAAAATAGCATCATGACCTTCAAG 271  
397 GGGGTGCAAAATAGCATCATGACCTTCAAG 427  
DB

RESULT 7  
AF529823 596 bp mRNA linear PRI 17-SEP-2002  
LOCUS Homo sapiens clone Ramos 9 AID (AID) mRNA, partial cds.  
DEFINITION  
ACCESSION AF529823  
VERSION AF529823.1 GI:22297233  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Martin,A. and Scharff,M.D.  
TITLE Somatic hypermutation of the AID transgene in B and non-B cells  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12304-12308 (2002)  
REFERENCE  
AUTHORS Martin,A. and Scharff,M.D.  
TITLE Direct Submmission  
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of  
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA  
FEATURES  
Location/Qualifiers  
source  
1. .>596  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="Ramos 9"  
1. .>596  
/gene="AID"  
1. .>596  
/gene="AID"  
/note="Integrated into Burkitt's lymphoma cell line Ramos"  
/codon\_start=1

gene  
1. .>596  
/gene="AID"  
CDS  
1. .>596  
/gene="AID"  
/note="Integrated into Burkitt's lymphoma cell line Ramos"  
/codon\_start=1  
/product="AID"  
/protein\_id="AAM95410.1"  
/db\_xref="GI:22297234"  
/translation="MDSLMMNRKFLYQKNVWAKGRRETYLCYVWRDASPSFL  
DFGLRNKNGCHEVLLFLRYISDMDDPGRCTRVTFWTFMSPTCDARHVADELGNP  
NLSLRIFTLARYFCEDRKAPBEGRLHRAQVQIALMTFKDYFCMNTVEVNERHRTFK  
AMEGLHENSVRLSRQLRRIILPLVEYDDLRDARFTLGL"



## ORIGIN

Query Match 100.0%; Score 271; DB 9; Length 596;  
Best Local Similarity 100.0%; Pred. No. 5.9e-50;  
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACGGCTGCCACGTGAATGCTCTTCCTCCGCTACATCTCGAGCTGGACCTAGACCTT 60  
DB 157 AACGGCTGCCACGTGAATGCTCTTCCTCCGCTACATCTCGAGCTGGACCTAGACCTT 216  
QY 61 GGGCGCTGCTACCGCGCTACCTGTTTCACTCTGGAGCCCTGCTACGACTGTGCCGA 120  
DB 217 GGGCGCTGCTACCGCGCTACCTGTTTCACTCTGGAGCCCTGCTACGACTGTGCCGA 276  
QY 121 CATGTGGCCGACCTTTCTGCGAGGGAACCCCACTTCACTGAGATCTTCAACCGCGGC 180  
DB 277 CATGTGGCCGACCTTTCTGCGAGGGAACCCCACTTCACTGAGATCTTCAACCGCGGC 336  
QY 181 CTCTACTTCTGTAGGACCGCAAGCTGAGCCCGAGGCGCTGCGCGCTGACCGCGCC 240  
DB 337 CTCTACTTCTGTAGGACCGCAAGCTGAGCCCGAGGCGCTGCGCGCTGACCGCGCC 396  
QY 241 GGGGTGCAATAGCCATCATGACCTTCAAG 271  
DB 397 GGGGTGCAATAGCCATCATGACCTTCAAG 427

RESULT 8  
AF529825 596 bp mRNA linear PRI 17-SEP-2002

LOCUS Homo sapiens clone Ramos 11 A1D (A1D) mRNA, partial cds.  
DEFINITION AF529825  
ACCESSION AF529825.1 GI:22297237

VERSION AF529825.1 GI:22297237  
KEYWORDS Homo sapiens (human)  
SOURCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 596)  
AUTHORS Martin, A. and Scharff, M.D.

TITLE Somatic hypermutation of the A1D transgene in B and non-B cells  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12304-12308 (2002)

REFERENCE 2 (bases 1 to 596)  
AUTHORS Martin, A. and Scharff, M.D.

TITLE Direct Submission  
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of  
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA

FEATURES  
location/Qualifiers

source  
1..596  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="Ramos 11"  
1..>596  
/gene="A1D"  
1..>596  
/gene="A1D"  
/note="Integrated into Burkitt's lymphoma cell line Ramos"

CDS  
1..>596  
/codon\_start=1  
/product="A1D"  
/protein\_id="AAM95412.1"  
/db\_xref="GI:22297238"

ORIGIN  
Query Match 100.0%; Score 271; DB 9; Length 596;  
Best Local Similarity 100.0%; Pred. No. 5.9e-50;  
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACGGCTGCCACGTGAATGCTCTTCCTCCGCTACATCTCGAGCTGGACCTAGACCTT 60

DB 157 AACGGCTGCCACGTGAATGCTCTTCCTCCGCTACATCTCGAGCTGGACCTAGACCTT 216

QY 61 GGGCGCTGCTACCGCGCTACCTGTTTCACTCTGGAGCCCTGCTACGACTGTGCCGA 120

DB 217 GGGCGCTGCTACCGCGCTACCTGTTTCACTCTGGAGCCCTGCTACGACTGTGCCGA 276

QY 121 CATGTGGCCGACCTTTCTGCGAGGGAACCCCACTTCACTGAGATCTTCAACCGCGGC 180

DB 277 CATGTGGCCGACCTTTCTGCGAGGGAACCCCACTTCACTGAGATCTTCAACCGCGGC 336

QY 181 CTCTACTTCTGTAGGACCGCAAGCTGAGCCCGAGGCGCTGCGCGCTGACCGCGCC 240

DB 337 CTCTACTTCTGTAGGACCGCAAGCTGAGCCCGAGGCGCTGCGCGCTGACCGCGCC 396

QY 241 GGGGTGCAATAGCCATCATGACCTTCAAG 271

DB 397 GGGGTGCAATAGCCATCATGACCTTCAAG 427

RESULT 9  
AF529829 596 bp mRNA linear SYN 17-SEP-2002

LOCUS Mus musculus clone 2 transgenic Homo sapiens A1D (A1D) mRNA,  
DEFINITION AF529829 complete cds.  
ACCESSION AF529829  
VERSION AF529829.1 GI:22297245

KEYWORDS Mus musculus (house mouse)  
SOURCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 596)  
AUTHORS Martin, A. and Scharff, M.D.

TITLE Somatic hypermutation of the A1D transgene in B and non-B cells  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12304-12308 (2002)

REFERENCE 2 (bases 1 to 596)  
AUTHORS Martin, A. and Scharff, M.D.

TITLE Direct Submission  
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of  
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA

FEATURES  
location/Qualifiers

source  
1..596  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="2"  
/cell\_type="hybridoma P1-5"  
/transgenic  
1..596  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
1..>596  
/gene="A1D"  
1..>596  
/gene="A1D"  
/note="Integrated into Burkitt's lymphoma cell line Ramos"

CDS  
1..>596  
/codon\_start=1  
/product="A1D"  
/protein\_id="AAM95416.1"  
/db\_xref="GI:22297246"

ORIGIN  
Query Match 100.0%; Score 271; DB 12; Length 596;  
Best Local Similarity 100.0%; Pred. No. 5.9e-50;  
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACGGCTGCCACGTGAATGCTCTTCCTCCGCTACATCTCGAGCTGGACCTAGACCTT 60



Db 157 AACGCTGCCACGTGGAATTGCTCTTCCTCCGCTACATTCGGACTGGAGCCTAGACCT 216

QY 61 GGCGGCTGTACACGCGCTACCTGATTCACCTCTCGAGGCCCTGTACGACTGTGCGCGA 120

Db 217 GGCGGCTGTACACGCGCTACCTGATTCACCTCTCGAGGCCCTGTACGACTGTGCGCGA 276

QY 121 CATGTGCGCGACTTTCTGTGCGAGGAAACCCCACTCAGTCTGAGATCTTCAACCGCGCG 180

Db 277 CATGTGCGCGACTTTCTGTGCGAGGAAACCCCACTCAGTCTGAGATCTTCAACCGCGCG 336

QY 181 CTCTACTTCTGTGAGACCGCGAAGCTGAGCGGCTGTGCGGCTGTGACCGCGCC 240

Db 337 CTCTACTTCTGTGAGACCGCGAAGCTGAGCGGCTGTGCGGCTGTGACCGCGCC 396

QY 241 GGGGTGCAATAGGCATCATGACCTTCAAG 271

Db 397 GGGGTGCAATAGGCATCATGACCTTCAAG 427

RESULT 10  
AF529831 596 bp mRNA linear SYN 17-SEP-2002  
LOCUS AF529831  
DEFINITION Mus musculus clone 4 transgenic Homo sapiens AID (AID) mRNA,  
complete cds.  
ACCESSION AF529831  
VERSION AF529831.1 GI:22297249  
KEYWORDS  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
AUTHORS Martin, A. and Scharff, M.D.  
TITLE Somatic hypermutation of the AID transgene in B and non-B cells  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12304-12308 (2002)  
AUTHORS Martin, A. and Scharff, M.D.  
TITLE Direct Submission  
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of  
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA

FEATURES  
source  
1..596  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="5"  
/cell\_type="hybridoma P1-5"  
/transgenic  
1..596  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
1..>596  
/gene="AID"  
1..>596  
/gene="AID"  
/codon\_start=1  
/product="AID"  
/protein\_id="AA95418.1"  
/db\_xref="GI:22297250"  
/translation="MDSLMMRRKFLYQFNKVMKGRRETYLCYVVKRSDATSPSL  
DFGVLNKGCHVELFLRYISDWDLDPRGCRVYVFTSMSPCYDCARHVAFLKGNP  
NLSIRFTALVFCEDRKAPBGLRLHRAVQIALMTFKDYFCMNTFVENHRTFK  
AMEGLHNSVRLSRQFRRLILPLYEVDLDLDAFRTLGL"

ORIGIN  
Query Match 100.0%; Score 271; DB 12; Length 596;  
Best Local Similarity 100.0%; Pred. No. 5.9e-50;  
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACGGCTGCCACGTGGAATTGCTCTTCCTCCGCTACATTCGGACTGGAGCCTAGACCT 60  
Db 157 AACGGCTGCCACGTGGAATTGCTCTTCCTCCGCTACATTCGGACTGGAGCCTAGACCT 216

QY 61 GGCGGCTGTACACGCGCTACCTGATTCACCTCTCGAGGCCCTGTACGACTGTGCGCGA 120

Db 217 GGCGGCTGTACACGCGCTACCTGATTCACCTCTCGAGGCCCTGTACGACTGTGCGCGA 276

QY 121 CATGTGCGCGACTTTCTGTGCGAGGAAACCCCACTCAGTCTGAGATCTTCAACCGCGCG 180

Db 277 CATGTGCGCGACTTTCTGTGCGAGGAAACCCCACTCAGTCTGAGATCTTCAACCGCGCG 336

QY 181 CTCTACTTCTGTGAGACCGCGAAGCTGAGCGGCTGTGCGGCTGTGACCGCGCC 240

Db 337 CTCTACTTCTGTGAGACCGCGAAGCTGAGCGGCTGTGCGGCTGTGACCGCGCC 396

QY 241 GGGGTGCAATAGGCATCATGACCTTCAAG 271

Db 397 GGGGTGCAATAGGCATCATGACCTTCAAG 427

RESULT 11  
AF529832 596 bp mRNA linear SYN 17-SEP-2002  
LOCUS AF529832  
DEFINITION Mus musculus clone 5 transgenic Homo sapiens AID (AID) mRNA,  
complete cds.  
ACCESSION AF529832  
VERSION AF529832.1 GI:22297251  
KEYWORDS  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
AUTHORS Martin, A. and Scharff, M.D.  
TITLE Somatic hypermutation of the AID transgene in B and non-B cells  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12304-12308 (2002)  
AUTHORS Martin, A. and Scharff, M.D.  
TITLE Direct Submission  
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of  
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA

FEATURES  
source  
1..596  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="5"  
/cell\_type="hybridoma P1-5"  
/transgenic  
1..596  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
1..>596  
/gene="AID"  
1..>596  
/gene="AID"  
/codon\_start=1  
/product="AID"  
/protein\_id="AA95419.1"  
/db\_xref="GI:22297252"  
/translation="MDSLMMRRKFLYQFNKVMKGRRETYLCYVVKRSDATSPSL  
DFGVLNKGCHVELFLRYISDWDLDPRGCRVYVFTSMSPCYDCARHVAFLKGNP  
NLSIRFTALVFCEDRKAPBGLRLHRAVQIALMTFKDYFCMNTFVENHRTFK  
AMEGLHNSVRLSRQFRRLILPLYEVDLDLDAFRTLGL"

ORIGIN  
Query Match 100.0%; Score 271; DB 12; Length 596;  
Best Local Similarity 100.0%; Pred. No. 5.9e-50;  
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACGGCTGCCACGTGGAATTGCTCTTCCTCCGCTACATTCGGACTGGAGCCTAGACCT 60  
Db 157 AACGGCTGCCACGTGGAATTGCTCTTCCTCCGCTACATTCGGACTGGAGCCTAGACCT 216



QY 61 GGCCGCTGCTACCGCGCTGCTACCTCTCTGAGCCCTGCTGCTACGACTGTGCCCCGA 120  
DB 217 GGCCGCTGCTACCGCGCTGCTACCTCTCTGAGCCCTGCTGCTACGACTGTGCCCCGA 276  
QY 121 CATGTGGCGGACCTTTCTGCGAGAGGAAACCCCACTTCAGTCTGAGGATCTTCAACCGGCGC 180  
DB 277 CATGTGGCGGACCTTTCTGCGAGAGGAAACCCCACTTCAGTCTGAGGATCTTCAACCGGCGC 336  
QY 181 CTCTACTTCTGTGAGGACCGCAAGGCTGAGCCCGAGGGGCTCGGCGGCTGACCGCGCC 240  
DB 337 CTCTACTTCTGTGAGGACCGCAAGGCTGAGCCCGAGGGGCTCGGCGGCTGACCGCGCC 396  
QY 241 GGCGTGCAATAGCCATCATGACCTTCAAG 271  
DB 397 GGCGTGCAATAGCCATCATGACCTTCAAG 427

RESULT 12  
AF529833 596 bp mRNA linear SYN 17-SEP-2002  
LOCUS Mus musculus clone 6 transgenic Homo sapiens AID (AID) mRNA,  
DEFINITION complete cds.  
ACCESSION AF529833  
VERSION AF529833.1 GI:22297253  
KEYWORDS  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
AUTHORS Martin, A. and Scharff, M.D.  
TITLE Somatic hypermutation of the AID transgene in B and non-B cells  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12304-12308 (2002)  
REVIEW 2 (bases 1 to 596)  
AUTHORS Martin, A. and Scharff, M.D.  
TITLE Direct Submision  
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of  
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA

FEATURES  
source  
1..596  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="6"  
/cell\_type="hybridoma PI-5"  
/transgenic  
1..596  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
1..>596  
/gene="AID"  
1..>596  
/gene="AID"  
/codon\_start=1  
/product="AID"  
/protein\_id="AA095420.1"  
/db\_xref="GI:22297253"  
/translation="MDSLMNRKFLYQKVMKAGRETYVYVYKRRDSATSFSL  
DFGLRNKNGCHELLFLRYISMDLDPGRCYAVTWFTSWSPCYDCAHVADFLRGNP  
NLSIRITFALVYFEDRKAPDEGLRLHRAVGQIALMTFKDYFYCMNTEVENHRTFK  
AMEGLHNSVRLSRQLRILLPLYEVDLDARFTLGL"

ORIGIN  
Query Match 100.0%; Score 271; DB 12; Length 596;  
Best Local Similarity 100.0%; Pred. No. 5.9e-50;  
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AAGCGCTGCAACGCGAATGCTCTCTCGCTACATCTCGGACTGGAGCTTGAACCT 60  
DB 157 AAGCGCTGCAACGCGAATGCTCTCTCGCTACATCTCGGACTGGAGCTTGAACCT 216  
QY 61 GGCGCTGCTACCGCGCTGCTACCTCTCTGAGCCCTGCTGCTACGACTGTGCCCCGA 120

DB 217 GGCCGCTGCTACCGCGCTGCTACCTCTCTGAGCCCTGCTGCTACGACTGTGCCCCGA 276  
QY 121 CATGTGGCGGACCTTTCTGCGAGAGGAAACCCCACTTCAGTCTGAGGATCTTCAACCGGCGC 180  
DB 277 CATGTGGCGGACCTTTCTGCGAGAGGAAACCCCACTTCAGTCTGAGGATCTTCAACCGGCGC 336  
QY 181 CTCTACTTCTGTGAGGACCGCAAGGCTGAGCCCGAGGGGCTCGGCGGCTGACCGCGCC 240  
DB 337 CTCTACTTCTGTGAGGACCGCAAGGCTGAGCCCGAGGGGCTCGGCGGCTGACCGCGCC 396  
QY 241 GGCGTGCAATAGCCATCATGACCTTCAAG 271  
DB 397 GGCGTGCAATAGCCATCATGACCTTCAAG 427

RESULT 13  
AF529835 596 bp mRNA linear SYN 17-SEP-2002  
LOCUS Mus musculus clone 8 transgenic Homo sapiens AID (AID) mRNA,  
DEFINITION complete cds.  
ACCESSION AF529835  
VERSION AF529835.1 GI:22297257  
KEYWORDS  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
AUTHORS Martin, A. and Scharff, M.D.  
TITLE Somatic hypermutation of the AID transgene in B and non-B cells  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12304-12308 (2002)  
REVIEW 2 (bases 1 to 596)  
AUTHORS Martin, A. and Scharff, M.D.  
TITLE Direct Submision  
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of  
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA

FEATURES  
source  
1..596  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="8"  
/cell\_type="hybridoma PI-5"  
/transgenic  
1..596  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
1..>596  
/gene="AID"  
1..>596  
/gene="AID"  
/codon\_start=1  
/product="AID"  
/protein\_id="AA095422.1"  
/db\_xref="GI:22297258"  
/translation="MDSLMNRKFLYQKVMKAGRETYVYVYKRRDSATSFSL  
DFGLRNKNGCHELLFLRYISMDLDPGRCYAVTWFTSWSPCYDCAHVADFLRGNP  
NLSIRITFALVYFEDRKAPDEGLRLHRAVGQIALMTFKDYFYCMNTEVENHRTFK  
AMEGLHNSVRLSRQLRILLPLYEVDLDARFTLGL"

ORIGIN  
Query Match 100.0%; Score 271; DB 12; Length 596;  
Best Local Similarity 100.0%; Pred. No. 5.9e-50;  
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AAGCGCTGCAACGCGAATGCTCTCTCGCTACATCTCGGACTGGAGCTTGAACCT 60  
DB 157 AAGCGCTGCAACGCGAATGCTCTCTCGCTACATCTCGGACTGGAGCTTGAACCT 216  
QY 61 GGCGCTGCTACCGCGCTGCTACCTCTCTGAGCCCTGCTGCTACGACTGTGCCCCGA 120



Db 217 GGGCGCTGCTACCGCGCTCACCTGATCCTCTGAGAGCCCTGCTACAGACTGTGCGCGA 276

QY 121 CATGTGGCCGACCTTTCTGCGAGGAAACCCCACTGAGTGTGAGATCTTCAACCGCGCC 180

Db 277 CATGTGGCCGACCTTTCTGCGAGGAAACCCCACTGAGTGTGAGATCTTCAACCGCGCC 336

QY 181 CTCTACTTCTGTGAGACCGCGAAGGCTGAGCCCGAGGGGCTGGCGGCTGTGACCGCGCC 240

Db 337 CTCTACTTCTGTGAGACCGCGAAGGCTGAGCCCGAGGGGCTGGCGGCTGTGACCGCGCC 396

QY 241 GGGGTGCAATAGCATCATGACCTTCAAG 271

Db 397 GGGGTGCAATAGCATCATGACCTTCAAG 427

RESULT 14

AF529836 596 bp mRNA linear SYN 17-SEP-2002

LOCUS AF529836

DEFINITION Mus musculus clone 9 transgenic Homo sapiens AID (AID) mRNA,

complete cds.

ACCESSION AF529836

VERSION AF529836.1 GI:22297259

KEYWORDS

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 596)

AUTHORS Martin,A. and Scharff,M.D.

TITLE Somatic hypermutation of the AID transgene in B and non-B cells

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12304-12308 (2002)

AUTHORS Martin,A. and Scharff,M.D.

TITLE Direct Submission

JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave, Bronx, NY 10461, USA

FEATURES

source

1..596

/organism="Mus musculus"

/mol\_type="mRNA"

/db\_xref="taxon:10090"

/clone="9"

/cell\_type="hybridoma PL-5"

/transgenic

1..596

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

1..>596

/gene="AID"

1..>596

/gene="AID"

/codon\_start=1

/product="AID"

/protein\_id="AA05423.1"

/db\_xref="GI:22297260"

/translation="MDSILMRKFLYQKNTNRKSKRRRETYLCVYTRRSATSFSL  
DPEYLRNNGCHEVELFLRYISDWIDPGRCYRVWFTWSMPCYCAHVAFLRGNP  
NLSLRFTARLYFCEDRKAPBGLRLHRAQVQIAIMTKDYFCWNTFVNHERTEK  
AMEGLHENSVRLSRQLRRLILPLEYVDLDRDAFRTGL"

ORIGIN

Query Match 100.0%; Score 271; DB 12; Length 596;

Best Local Similarity 100.0%; Pred. No. 5,9e-50;

Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCGTGCACGCTGAATTTGCTCTCCCGCTACATCTCGAGCTGGAGCTAGACCT 60

Db 157 AAGCGTGCACGCTGAATTTGCTCTCCCGCTACATCTCGAGCTGGAGCTAGACCT 216

QY 61 GGGCGTGTCTACCGCGCTCACTGTTCACCTCTCTGAGAGCCCGTGTACAGCTGTGCCGA 120

Db 217 GGGCGTGTCTACCGCGCTCACTGTTCACCTCTCTGAGAGCCCGTGTACAGCTGTGCCGA 276

QY 121 CATGTGGCCGACCTTTCTGCGAGGAAACCCCACTGAGTGTGAGATCTTCAACCGCGCC 180

Db 277 CATGTGGCCGACCTTTCTGCGAGGAAACCCCACTGAGTGTGAGATCTTCAACCGCGCC 336

QY 181 CTCTACTTCTGTGAGACCGCGAAGGCTGAGCCCGAGGGGCTGGCGGCTGTGACCGCGCC 240

Db 337 CTCTACTTCTGTGAGACCGCGAAGGCTGAGCCCGAGGGGCTGGCGGCTGTGACCGCGCC 396

QY 241 GGGGTGCAATAGCATCATGACCTTCAAG 271

Db 397 GGGGTGCAATAGCATCATGACCTTCAAG 427

RESULT 15

BT007402 597 bp mRNA linear PRI 13-MAY-2003

LOCUS BT007402

DEFINITION Homo sapiens activation-induced cytidine deaminase mRNA, complete

cds.

ACCESSION BT007402

VERSION BT007402.1 GI:30583642

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 597)

AUTHORS Kalline,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S., Koundinya,M., Raphael,J., Moreira,D., Kelley,T., Labaer,J., Lin,Y., Phelan,M. and Farmer,A.

TITLE Cloning of human full-length CDSs in BD Creator(TM) System Donor vector

JOURNAL Unpublished

AUTHORS 2 (bases 1 to 597)

Kalline,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S., Koundinya,M., Raphael,J., Moreira,D., Kelley,T., Labaer,J., Lin,Y., Phelan,M. and Farmer,A.

TITLE Direct Submission

JOURNAL Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow Circle, Palo Alto, CA 94303, USA

COMMENT This CDS clone is a part of a collection of human full length expression clones generated by BD Biosciences Clontech and the Harvard Institute of Proteomics. Each CDS has been cloned in two forms: with and without stop-codon (to allow fusion with C-terminal tag). The CDS has been directionally cloned using BD In-Fusion(TM) cloning system between the SalI and HindIII sites of the pDNR-DUAL vector. Additional sequences in the clone: 'ACC' after SalI site and before 'ATG' to provide Kozak consensus sequence; 'GG' after last codon and before HindIII site to maintain reading frame. Clone distribution: <http://bioinfo.clontech.com/orfclones>.

FEATURES

source

1..597

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="GH00009X1.0"

/clone\_1lb="BD Creator(TM) CDS library derived from MGC collection"

/lab\_host="DH5alpha T1 resistant"

/notes="Vector: pDNR-Dual"

1..597

/codon\_start=1

/product="activation-induced cytidine deaminase"

/protein\_id="AAP36066.1"

/db\_xref="GI:30583643"

/translation="MDSILMRKFLYQKNTNRKSKRRRETYLCVYTRRSATSFSL  
DPEYLRNNGCHEVELFLRYISDWIDPGRCYRVWFTWSMPCYCAHVAFLRGNP  
NLSLRFTARLYFCEDRKAPBGLRLHRAQVQIAIMTKDYFCWNTFVNHERTEK  
AMEGLHENSVRLSRQLRRLILPLEYVDLDRDAFRTGL"

ORIGIN

Query Match 100.0%; Score 271; DB 9; Length 597;

Best Local Similarity 100.0%; Pred. No. 5,9e-50;



	Matches	271;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	1	AAAGCTGCCACGTGGAATTGCTCTCTCTCCGCTACATCTCGGACTGGGACCTAGACCT	60							
Db	157	AAAGCTGCCACGTGGAATTGCTCTCTCTCCGCTACATCTCGGACTGGGACCTAGACCT	216							
Qy	61	GGCGCTGCTACCGGCTCACTGTTCACTCTCGAGCCCTGCTACGACTGTGCCGA	120							
Db	217	GGCGCTGCTACCGGCTCACTGTTCACTCTCGAGCCCTGCTACGACTGTGCCGA	276							
Qy	121	CATGTGGCCGACTTTCTGCGAGGGAACCCCAACTCATGTGAGATCTTCAACGGCGC	180							
Db	277	CATGTGGCCGACTTTCTGCGAGGGAACCCCAACTCATGTGAGATCTTCAACGGCGC	336							
Qy	181	CTCTACTTCTGTAGAGACCGCAAGGCTGAGCCCGAGGGCTGGGGCGCTGCACCGCGC	240							
Db	337	CTCTACTTCTGTAGAGACCGCAAGGCTGAGCCCGAGGGCTGGGGCGCTGCACCGCGC	396							
Qy	241	GGGTCGCAATAGCCATCATGACCTTCAAAG	271							
Db	397	GGGTCGCAATAGCCATCATGACCTTCAAAG	427							

Search completed: March 12, 2004, 21:22:31  
Job time : 854.181 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 12, 2004, 12:26:39 ; Search time 108.517 Seconds

(without alignments)  
10609.014 Million cell updates/sec

Title: US-09-966-880A-13

Perfect score: 271  
Sequence: 1 aacgcctgcacacgtgaatt.....agcatcatgaccttcaag 271

Scoring table: IDENTITY NTC  
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_29Jan04:\*  
1: geneseqn180s:\*  
2: geneseqn190s:\*  
3: geneseqn200s:\*  
4: geneseqn2001s:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002s:\*  
7: geneseqn2003as:\*  
8: geneseqn2003bs:\*  
9: geneseqn2003cs:\*  
10: geneseqn2004s:\*  
  
Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	271	100.0	271	3 AAC55317	AAC55317 Human act
2	271	100.0	597	8 ADB88952	ADB88952 AID gene
3	271	100.0	1543	7 ABX05468	ABX05468 Human nov
4	271	100.0	2791	6 AB873287	AB873287 DNA encod
5	271	100.0	2791	6 AB873288	AB873288 DNA encod
6	271	100.0	2818	3 AAC55312	AAC55312 Human act
7	271	100.0	6564	3 AAC55314	AAC55314 Human act
8	271	100.0	11204	3 AAC55339	AAC55339 Human act
9	271	100.0	11204	6 AB873286	AB873286 DNA encod
10	221.6	81.8	2440	3 AAC55307	AAC55307 Mouse act
11	87.4	32.3	201143	2 ABR63568	ABR63568 Human DNA
12	85.8	31.7	610	2 AA472058	AA472058 CDNA enco
13	85.8	31.7	610	3 AA472058	AA472058 CDNA enco
14	85.8	31.7	819	4 AA472058	AA472058 CDNA enco
15	85.8	31.7	944	6 ABR199876	ABR199876 Human sec
16	85.8	31.7	950	4 AA41420	AA41420 CDNA enco
17	85.8	31.7	987	4 AA416264	AA416264 Human pro
18	85.8	31.7	987	4 AA416264	AA416264 CDNA enco
19	85.8	31.7	1120	4 AA159847	AA159847 Human pol
20	85.8	31.7	1143	4 AA158061	AA158061 Human pol
21	85.8	31.7	1638	6 AB061187	AB061187 Human pol
22	85.8	31.7	6131	4 AA42187	AA42187 Genomic s
23	85.8	31.7	6131	4 AA42187	AA42187 Human imm

24	85	31.4	1534	2 AA20856	AA20856 Polynucle
25	85	31.4	1534	4 AA559293	AA559293 Human CDN
26	85	31.4	1534	6 ABA90962	ABA90962 Human pol
27	85	31.4	1717	9 ADD18987	ADD18987 Human dis
28	85	31.4	2151	6 AA652572	AA652572 CDNA sequ
29	83.4	30.8	1446	3 AA12409	AA12409 CDNA enco
30	82.6	30.5	1055	6 AAD24392	AAD24392 Human RNA
31	78.4	29.0	122748	6 ABT10719	ABT10719 Human bre
32	78.4	28.9	371	8 ACH43025	ACH43025 Human foe
33	77.4	28.6	1526	9 ADD18895	ADD18895 Human dis
34	76.8	28.3	823	9 ADE25658	ADE25658 Human CDN
35	76.8	28.3	1348	6 ABR6785	ABR6785 Gene #328
36	73.4	27.1	1567	4 AA193080	AA193080 Human pol
37	70.2	25.9	716	2 AA215848	AA215848 Human gen
38	68.4	25.2	2338	4 AA117654	AA117654 Human CDN
39	68.4	25.2	2773	9 ADB65605	ADB65605 Human CDN
40	64.8	23.9	572	4 AA118764	AA118764 Probe #86
41	64.8	23.9	572	4 ABA63759	ABA63759 Human foe
42	64.8	23.9	572	4 AA143877	AA143877 Probe #12
43	64.8	23.9	572	4 ABA30950	ABA30950 Probe #94
44	64.8	23.9	572	4 AA37995	AA37995 Human bor
45	64.8	23.9	572	4 AA12273	AA12273 Human bra

ALIGNMENTS

RESULT 1  
AAC55317  
ID AAC55317 standard; DNA; 271 BP.  
XX AC AAC55317;  
DT 05-FEB-2001 (first entry)  
DE Human activation-induced cytidine deaminase exon 3 SEQ ID NO:13.  
XX  
KM Activation-induced cytidine deaminase; AID; cytidine deaminase;  
KM immune related disease; allergy; allergic disease; anti-allergic;  
KM anti-inflammatory; antiasthmatic; ophthalmological; anti-HIV; dermatological;  
KM gene therapy; B cell associated immune system disorder; food allergy;  
KM immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
KM IgA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;  
KM drug allergy; allergic rhinitis; Rosen disease; discoid dermatitis; AIDS;  
KM ataxia telangiectasia; common variable immunodeficiency disorder;  
KM major histocompatibility class II deficiency disease;  
KM auto immunodeficiency syndrome; Igs subclass selection disorder; ds.  
XX  
OS Homo sapiens.  
XX  
FN WO200058480-A1.  
XX  
PD 05-OCT-2000.  
XX  
PF 28-MAR-2000; 2000WO-JP001918.  
XX  
PR 29-MAR-1999; 99JP-00087192.  
XX  
PR 24-JUN-1999; 99JP-00178999.  
XX  
PR 27-DEC-1999; 99JP-00371382.  
XX  
PA (NTSB) JAPAN TOBACCO INC.  
XX  
PA (HONJ) HONJO T.  
XX  
PI Honjo T, Muramatsu M;  
XX  
WP: 2000-611715/58.  
XX  
PT Nucleic acid encoding activation induced cytidine deaminase, useful as a  
XX target for drug development for immune-related diseases including  
XX allergies.  
XX  
PS Claim 18; Page 151; 174pp: Japanese.  
XX



CC The present invention describes an activation-induced cytidine deaminase  
CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and has  
CC cytidine activity similar to APOBEC-1. AID has anti-allergic, anti-anemic,  
CC anti-asthmatic, ophthalmological, anti-HIV and dermatological activities,  
CC and can be used in gene therapy. AID polynucleotides are useful in  
CC methods for identifying drugs for the treatment of B cell associated  
CC immune system disorders, immunodeficiency diseases and allergies, such as  
CC immunoglobulin A (IgA) deficiency disease, IgA nephritis, gamma-  
CC globulinemia, atopic dermatitis, allergic colitis, asthma, food allergy,  
CC drug allergy, allergic rhinitis, Rosen disease, DiGeorge disease, ataxia  
CC telangiectasia, common variable immunodeficiency disorder, MHC (major  
CC histocompatibility class II deficiency disease, AIDS (auto  
CC immunodeficiency syndrome), elevated IgE disorder, and IgG subclass  
CC selection disorder. The DNA sequences encoding AID may be used for gene  
CC therapy and the antibodies to the AID protein may be used for diagnosis  
CC and treatment of these disorders. The present sequence represents the  
CC exon 3 genomic DNA sequence of human AID  
CC  
XX

SO Sequence 271 BP; 47 A; 95 C; 76 G; 53 T; 0 U; 0 Other;

Query Match 100.0%; Score 271; DB 3; Length 271;

Best Local Similarity 100.0%; Pred.No. 3.3e-58;

Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACGGCTGCGACGTGGAAATGCTTCTTCCCTGCTACATCTGGAGCTGGAGCTAGACCT 60  
DB 1 AACGGCTGCGACGTGGAAATGCTTCTTCCCTGCTACATCTGGAGCTGGAGCTAGACCT 60  
QY 61 GGGCGGTGTACCGGGTCACTGTTCACTCTGAGAGCCCTGTACAGCTGAGCCCGA 120  
DB 61 GGGCGGTGTACCGGGTCACTGTTCACTCTGAGAGCCCTGTACAGCTGAGCCCGA 120  
QY 121 CATGTGCGCGACTTCTGCGAGAGAACCCCACTCATGTGAGGATCTTCAACCGCGCGC 180  
DB 121 CATGTGCGCGACTTCTGCGAGAGAACCCCACTCATGTGAGGATCTTCAACCGCGCGC 180  
QY 181 CTCTACTTCTGTGAGAGACCGCAAGGCTGAGCCCGAGGGGCTGCGGCTGACCGCGCC 240  
DB 181 CTCTACTTCTGTGAGAGACCGCAAGGCTGAGCCCGAGGGGCTGCGGCTGACCGCGCC 240  
QY 241 GGGGTGCAATAGCCATCATGACCTTCAAG 271  
DB 241 GGGGTGCAATAGCCATCATGACCTTCAAG 271

RESULT 2

ID ADB88952 standard; DNA; 597 BP.

AC ADB88952;

XX 04-DEC-2003 (first entry)

DE AID gene as substrate used in AID mutagenic method.

XX ds; cytosstatic; virucide; antidote; mutation;

KM activation-induced cytidine deaminase; cell phenotype;

KM monoclonal antibody; antigen; cross-reactivity; tumour;

XX germ warfare agent; toxin; V region.

OS Homo sapiens.

XX Key Location/Qualifiers

FT mutation /replace(9,A)

FT mutation /tag= a /note= "mutation in PI-5 hybridoma cell"

FT mutation /replace(12,A)

FT mutation /tag= b /note= "mutation in Ramos cell"

FT mutation /replace(14,C)

FT mutation /tag= c /note= "mutation in CHO cell"

FT mutation /tag= d /note= "mutation in CHO cell"

FT mutation /tag= e /note= "mutation in PI-5 hybridoma cell"

FT mutation /tag= f /note= "mutation in PI-5 hybridoma cell"

FT mutation /tag= g /note= "mutation in Ramos cell"

FT mutation /tag= h /note= "mutation in PI-5 hybridoma cell"

FT mutation /tag= i /note= "mutation in Ramos cell"

FT mutation /tag= j /note= "mutation in PI-5 hybridoma cell"

FT mutation /tag= k /note= "mutation in PI-5 hybridoma cell"

FT mutation /tag= l /note= "mutation in PI-5 hybridoma cell"

FT mutation /tag= m /note= "mutation in Ramos cell"

FT mutation /tag= n /note= "mutation in Ramos cell"

FT mutation /tag= o /note= "mutation in CHO cell"

FT mutation /tag= p /note= "mutation in Ramos cell"

FT mutation /tag= q /note= "mutation in Ramos cell"

FT mutation /tag= r /note= "mutation in Ramos cell"

FT mutation /tag= s /note= "mutation in Ramos cell"

FT mutation /tag= t /note= "mutation in Ramos cell"

FT mutation /tag= u /note= "mutation in Ramos cell"

FT mutation /tag= v /note= "mutation in Ramos cell"

FT mutation /tag= w /note= "mutation in Ramos cell, G mutation in CHO cell"

FT mutation /tag= x /note= "mutation in Ramos cell, A mutation in CHO cell"

FT mutation /tag= y /note= "mutation in Ramos cell"

FT mutation /tag= z /note= "mutation in PI-5 hybridoma cell"

FT mutation /tag= aa /note= "mutation in CHO cell"

FT mutation /tag= ab /note= "mutation in CHO cell"



CC	producing mutated genes and proteins, determining the effect of a
CC	mutation on a protein or cell phenotype, and producing genes and proteins
CC	with altered properties. It is especially useful for producing antibodies
CC	with altered affinities or specificities for an antigen, or
CC	increased/decreased cross-reactivity for a second antigen, e.g. to
CC	produce antibodies with greater affinity/specificity useful
CC	therapeutically to treat tumours, viruses such as Ebola and Lassa fever
CC	or against germ warfare agents, toxins (e.g. ricin). The method may
CC	similarly be used to alter affinity, specificity or cross-reactivity of a
CC	monoclonal antibody. This sequence represents the AID gene used in the
CC	method of the invention to induce mutations in the sequence.
XX	
SQ	Sequence 597 BP; 128 A; 164 C; 155 G; 150 T; 0 U; 0 Other;
Query Match	100.0%; Score 271; DB 8; Length 597;
Best Local Similarity	100.0%; Pred. No. 3.6e-58;
Matches 271; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1 AACGGCTGCCACCGGGAATGTCTTCTCCTCCGCTACATCTCGACTGGAGACTTAAGCCT 60
DB	157 AACGGCTGCCACCGGGAATGTCTTCTCCTCCGCTACATCTCGACTGGAGACTTAAGCCT 216
QY	61 GGCGGCTGTCACCGCGTCACCTGGTTACCTCTCGAGACCCTTGACGACTGTGCCGA 120
DB	217 GGCGGCTGTCACCGCGTCACCTGGTTACCTCTCGAGACCCTCTACGACTGTGCCGA 276
QY	121 CANGTGCCGCACTTTCTGCAGAGGAAACCCCACTCATGCTGAGAGATTTCACCGCGCGC 180
DB	277 CATGTGGCGCACTTTCTGCAGAGGAAACCCCACTCATGCTGAGAGATTTCACCGCGCGC 336
QY	181 CTCTACTTTCTGTGAGAACCGCAAGCTGAGCCCGAGGGGCTGCGGCGGCTGCAACGGCGC 240
DB	337 CTCTACTTTCTGTGAGAACCGCAAGCTGAGCCCGAGGGGCTGCGGCGGCTGCAACGGCGC 396
QY	241 GGGGTGCAATAAGCATCATGACTCTTAAG 271
DB	397 GGGGTGCAATAAGCATCATGACTCTTAAG 427
RESULT 3	
ABX05468	
ID	ABX05468 standard; cDNA; 1543 BP.
XX	
AC	ABX05468;
XX	
DT	17-JAN-2003 (first entry)
XX	
DE	Human novel polynucleotide #483.
XX	
KW	Human; gene; ss; genetic disorder; gene mapping; medical imaging; cancer;
KW	neurodegenerative disorder; lymphoid cell disorder; osteoporosis;
KW	Parkinson's disease; Alzheimer's disease; bone degenerative disorder;
KW	osteoarthritis; periodontal disease; liver fibrosis; viral infection;
KW	fungal infection; bacterial infection; autoimmune disease; diabetes;
KW	atopic dermatitis.
XX	
OS	Homo sapiens.
XX	
EN	WO200274961-A1.
XX	
PD	26-SEP-2002.
XX	
PF	14-MAR-2002; 2002MO-USO05109.
XX	
PR	15-MAR-2001; 2001US-00810173.
PA	(HYSE-) HYSEQ INC.
PI	Tang YT, Zhou P, Goodrich R, Asundi V, Zhang J, Zhao QA, Ren F,
PI	Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M,
PI	Wehrman T, Wang J, Wang D, Drmanac RT;
XX	
RP	WPI; 2003-040556/03.



DR P-PSDB; ABU00390.  
XX  
XX New isolated polypeptides and polynucleotides, useful for preventing,  
PT treating or ameliorating medical conditions, such as cancer,  
PT neurodegenerative disorders, lymphoid cell disorders, bone degenerative  
PT disorders, and infections.  
XX  
XX Claim 1; SEQ ID NO 483; 235pp; English.  
XX  
XX The invention relates to human polynucleotides and the polypeptides they  
CC encode. The polynucleotides and polypeptides are useful in diagnostics,  
CC forensics, gene mapping, medical imaging, identification of mutations,  
CC responsible for genetic disorders or other traits, assessing biodiversity  
CC and producing many other types of data and products dependent on DNA and  
CC amino acid sequences. They are also useful for preventing, treating or  
CC ameliorating medical conditions, such as cancer, neurodegenerative  
CC disorders (e.g. Parkinson's disease, Alzheimer's disease), lymphoid cell  
CC disorders, osteoporosis, osteoarthritis, bone degenerative disorders,  
CC periodontal disease, liver fibrosis, infections (e.g. viral, fungal or  
CC bacterial) or autoimmune diseases (e.g. diabetes, atopic dermatitis).  
CC Sequences ABX04986-ABX0511 represent human polynucleotides of the  
CC invention. Note: The sequence data for this patent is not represented in  
CC the printed specification but is based on sequence information supplied  
CC by the European Patent Office  
XX  
SQ Sequence 1543 BP; 428 A; 334 C; 352 G; 429 T; 0 U; 0 Other;  
Query Match 100.0%; Score 271; DB 7; Length 1543;  
Best Local Similarity 100.0%; Pred. No. 4,2e-58;  
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AAGCGCTGCCAGTGGAAATGCTCTCTCCGCTACATCTCGAGCTGGACCTAGACCT 60  
DB 234 AAGCGCTGCCAGTGGAAATGCTCTCTCCGCTACATCTCGAGCTGGACCTAGACCT 293  
QY 61 GGGCGCTGACCGCGGACCGCTGCTCACTCTGAGGCCCGCTGACGACTGTGGCCGA 120  
DB 294 GGGCGCTGACCGCGGCTGCTCACTCTGAGGCCCGCTGACGACTGTGGCCGA 353  
QY 121 CATGTGGCCGACTTCTTGTGCGAGGAAACCCCACTGAGTCTGAGATCTTCAACCGCGCG 180  
DB 354 CATGTGGCCGACTTCTTGTGCGAGGAAACCCCACTGAGTCTGAGATCTTCAACCGCGCG 413  
QY 181 CTCTACTTCTGTGAGACCGCAAGGCTGAGGCCGCGGCTGCGCTGCAACCGCGCC 240  
DB 414 CTCTACTTCTGTGAGACCGCAAGGCTGAGGCCGCGGCTGCGCTGCAACCGCGCC 473  
QY 241 GGGGTGCAATATGACCATCATGACCTTCAAG 271  
DB 474 GGGGTGCAATATGACCATCATGACCTTCAAG 504

## RESULT 4

ABST73287  
ID ABST73287 standard; DNA; 2791 BP.

XX  
XX ABST73287;  
XX  
XX 04-DEC-2002 (first entry)  
XX  
XX DNA encoding human translocation del(12p) protein #2.  
XX  
XX Chromosome aberration; oncogenic fusion protein; cancer;  
XX proliferative disease; cellular protein isoform; heat shock protein 90;  
XX HSP-90; rheumatoid arthritis; cancer; haematopoietic disorder;  
XX T cell lymphoma; B cell lymphoma; chronic myeloid leukaemia; CML;  
XX acute myeloid leukaemia; AML; chronic myelomonocytic leukaemia; CMML;  
XX acute lymphoblastic leukaemia; ALL; APL; NHL; solid tumour;  
XX papillary thyroid carcinoma; Ewing's sarcoma; melanoma; liposarcoma;  
XX rhabdomyosarcoma; synovial sarcoma; viral infection; gene; ds.  
XX  
XX Homo sapiens.  
XX

PN W0200269900-A2.  
XX  
XX 12-SEP-2002.  
XX  
XX 01-MAR-2002; 2002MO-US006518.  
XX  
XX 01-MAR-2001; 2001US-0272751P.  
XX  
XX (CONF-) CONFORMA THERAPEUTICS CORP.  
XX  
XX Fritz LC, Burrows FJ;  
XX WPI; 2002-698710/75.  
XX  
XX P-PSDB; ABG95083.  
XX  
XX Treating genetically-defined disease associated with chromosomal  
PT aberrations yielding oncogenic fusion proteins, e.g. cell proliferative  
PT diseases, involves administering an inhibitor of heat shock protein 90.  
XX  
XX Disclosure; Page 246-247; 389pp; English.  
XX  
XX The invention describes a method of treating genetically-defined disease  
CC associated with chromosomal aberrations yielding oncogenic fusion  
CC proteins (I), treating cancerous cells containing (I) in a heterogeneous  
CC cell population, treating proliferative diseases associated with mutant  
CC protein or cellular protein isoforms (II) dependent on heat shock protein  
CC (HSP)-90, or selectively treating cells expressing (II) involving  
CC administering HSP90-inhibitor. The method is useful for treating  
CC genetically-defined disease with chromosomal aberration yielding  
CC oncogenic fusion protein, treating cancerous cells containing fusion  
CC protein in heterogeneous cell population, treating proliferative disease  
CC (e.g. rheumatoid arthritis or cancer) associated with mutant protein or  
CC cellular protein isoform dependent on heat shock protein (HSP)-90 (e.g.  
CC p53), or selectively treating cells expressing mutant protein or cellular  
CC protein isoform in a patient heterozygous for (II). The method is useful  
CC for treating a disease e.g. haematopoietic disorder such as T or B cell  
CC lymphoma, chronic myeloid leukaemia (CML), APL, ALL, NHL and CMML,  
CC or a disease characterised by a solid tumour such as papillary thyroid  
CC carcinoma, Ewing's sarcoma, melanoma, liposarcoma, rhabdomyosarcoma and  
CC synovial sarcoma. The method is also useful for treating viral  
CC infections. This represents the DNA sequence of a chromosome aberration  
SQ Sequence 2791 BP; 842 A; 548 C; 625 G; 776 T; 0 U; 0 Other;  
Query Match 100.0%; Score 271; DB 6; Length 2791;  
Best Local Similarity 100.0%; Pred. No. 4,5e-58;  
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AAGCGCTGCCAGTGGAAATGCTCTCTCCGCTACATCTCGAGCTGGACCTAGACCT 60  
DB 233 AAGCGCTGCCAGTGGAAATGCTCTCTCCGCTACATCTCGAGCTGGACCTAGACCT 292  
QY 61 GGGCGCTGACCGCGGCTGCTCACTCTGAGGCCCGCTGACGACTGTGGCCGA 120  
DB 293 GGGCGCTGACCGCGGCTGCTCACTCTGAGGCCCGCTGACGACTGTGGCCGA 352  
QY 121 CATGTGGCCGACTTCTTGTGCGAGGAAACCCCACTGAGTCTGAGATCTTCAACCGCGCG 180  
DB 353 CATGTGGCCGACTTCTTGTGCGAGGAAACCCCACTGAGTCTGAGATCTTCAACCGCGCG 412  
QY 181 CTCTACTTCTGTGAGACCGCAAGGCTGAGGCCGCGGCTGCGCTGCAACCGCGCC 240  
DB 413 CTCTACTTCTGTGAGACCGCAAGGCTGAGGCCGCGGCTGCGCTGCAACCGCGCC 472  
QY 241 GGGGTGCAATATGACCATCATGACCTTCAAG 271  
DB 473 GGGGTGCAATATGACCATCATGACCTTCAAG 503

## RESULT 5

ABST73288  
ID ABST73288 standard; DNA; 2791 BP.

XX  
XX



AC AAS73288;  
 XX  
 DT 04-DEC-2002 (first entry)  
 XX  
 DE DNA encoding human translocation del(12p) protein #3.  
 XX  
 KW Chromosome aberration; oncogenic fusion protein; cancer;  
 KW Proliferative disease; cellular protein isoform; heat shock protein 90;  
 KW HSP-90; rheumatoid arthritis; cancer; haematopoietic disorder;  
 KW T cell lymphoma; B cell lymphoma; chronic myeloid leukaemia; CML;  
 KW acute myeloid leukaemia; AML; chronic myelomonocytic leukaemia; CMML;  
 KW acute lymphoblastic leukaemia; ALL; APL; NHL; solid tumour;  
 KW papillary thyroid carcinoma; Ewing's sarcoma; melanoma; liposarcoma;  
 KW rhabdomyosarcoma; synovial sarcoma; viral infection; gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W0200269900-A2.  
 XX  
 PD 12-SEP-2002.  
 XX  
 PF 01-MAR-2002; 2002WO-US006518.  
 XX  
 PR 01-MAR-2001; 2001US-0272751P.  
 XX  
 PA (CONF-) CONFORMA THERAPEUTICS CORP.  
 XX  
 PI Fritz LC, Burrows FJ;  
 XX  
 DR WPI: 2002-698710/75.  
 DR P-PSDB; ABB95084.  
 XX  
 PT Treating genetically-defined disease associated with chromosomal  
 PT aberrations yielding oncogenic fusion proteins, e.g. cell proliferative  
 PT diseases, involves administering an inhibitor of heat shock protein 90.  
 XX  
 PS Disclosure; Page 248-249; 389pp; English.  
 XX  
 CC The invention describes a method of treating genetically-defined disease  
 CC associated with chromosomal aberrations yielding oncogenic fusion  
 CC proteins (I), treating cancerous cells containing (I) in a heterogeneous  
 CC cell population, treating proliferative diseases associated with mutant  
 CC protein or cellular protein isoforms (II) dependent on heat shock protein  
 CC (HSP)-90, or selectively treating cells expressing (II) involving  
 CC administering HSP90-inhibitor. The method is useful for treating  
 CC genetically-defined disease with chromosomal aberration yielding  
 CC oncogenic fusion protein, treating cancerous cells containing fusion  
 CC protein in heterogeneous cell population, treating proliferative disease  
 CC (e.g. rheumatoid arthritis or cancer) associated with mutant protein or  
 CC cellular protein isoform dependent on heat shock protein (HSP)-90 (e.g.  
 CC p53), or selectively treating cells expressing mutant protein or cellular  
 CC protein isoform in a patient heterozygous for (II). The method is useful  
 CC for treating a disease e.g. haematopoietic disorder such as T or B cell  
 CC lymphoma, chronic myeloid leukaemia (CML), APL, ALL, AML, NHL and CMML,  
 CC or a disease characterised by a solid tumour such as papillary thyroid  
 CC carcinoma, Ewing's sarcoma, melanoma, liposarcoma, rhabdomyosarcoma and  
 CC synovial sarcoma. The method is also useful for treating viral  
 CC infections. This represents the DNA sequence of a chromosome aberration  
 XX  
 SO Sequence 2791 BP; 842 A; 548 C; 625 G; 776 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 271; DB 6; Length 2791;  
 Best Local Similarity 100.0%; Pred. No. 4.5e-58;  
 Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 CATGTGGCCGACTTTCTGCGAGGAGACCCCACTCAGTCTGAGAACTTTGACCGCGCC 180  
 DB 353 CATGTGGCCGACTTTCTGCGAGGAGACCCCACTCAGTCTGAGAACTTTGACCGCGCC 412  
 QY 181 CTCTACTTCTGTGAGGACCGCAGAGGCTGAGGGGCTGCGGCGGCTGCACCGCGCC 240  
 DB 413 CTCTACTTCTGTGAGGACCGCAGAGGCTGAGGGGCTGCGGCGGCTGCACCGCGCC 472  
 QY 241 GGGGTGCAAAATAGCATCATGACCTTCAAG 271  
 DB 473 GGGGTGCAAAATAGCATCATGACCTTCAAG 503  
 RESULT 6  
 AAC55312  
 ID AAC55312 standard; cDNA; 2818 BP.  
 XX  
 AC AAC55312;  
 XX  
 DT 05-FEB-2001 (first entry)  
 XX  
 DE Human activation-induced cytidine deaminase encoding cDNA SEQ ID NO:7.  
 XX  
 KW Activation-induced cytidine deaminase; AID; cytidine deaminase;  
 KW immune related disease; allergy; allergic disease; anti-allergic;  
 KW antianemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;  
 KW gene therapy; B cell associated immune system disorder; food allergy;  
 KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
 KW IGA nephritis; gamma-globulinemia; atopic dermatitis; allergic colitis;  
 KW drug allergy; allergic rhinitis; Rosen disease; Digeorge disease; AIDS;  
 KW ataxia telangiectasia; common variable immunodeficiency disorder;  
 KW major histocompatibility class II deficiency disease;  
 KW auto immunodeficiency syndrome; Igg subclass selection disorder; ss.  
 KW  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 80..676  
 FT /\*tag= a  
 FT /product= "activation-induced cytidine deaminase"  
 PN W0200058480-A1.  
 XX  
 PD 05-OCT-2000.  
 XX  
 PF 28-MAR-2000; 2000WO-JP001918.  
 XX  
 PR 29-MAR-1999; 99JP-00087192.  
 PR 24-JUN-1999; 99JP-00178999.  
 PR 27-DEC-1999; 99JP-00371382.  
 XX  
 PA (NIBS) JAPAN TOBACCO INC.  
 PA (HONJ/) HONJYO T.  
 XX  
 PI Honjo T, Muramatsu M;  
 XX  
 DR WPI: 2000-611715/58.  
 DR P-PSDB; AAB24198.  
 XX  
 PT Nucleic acid encoding activation induced cytidine deaminase, useful as a  
 PT target for drug development for immune-related diseases including  
 PT allergies.  
 XX  
 PS Claim 3; Page 135-139; 174pp; Japanese.  
 XX  
 CC The present sequence encodes human activation-induced cytidine deaminase  
 CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and has  
 CC cytidine activity similar to APOBEC-1. AID has antiallergic, antianemic,  
 CC antiasthmatic, ophthalmological, anti-HIV and dermatological activities,  
 CC and can be used in gene therapy. AID polynucleotides are useful in  
 CC methods for identifying drugs for the treatment of B cell associated  
 CC immune system disorders, immunodeficiency diseases and allergies, such as  
 CC immunoglobulin A (IGA) deficiency disease, IGA nephritis, gamma-



CC globulinemia, atopic dermatitis, allergic colitis, asthma, food allergy,  
CC drug allergy, allergic rhinitis, Rosen disease, DiGeorge disease, ataxia  
CC telangiectasia, common variable immunodeficiency disorder, MHC (major  
CC histocompatibility class II deficiency disease, AIDS (auto  
CC immunodeficiency syndrome), elevated IgE disorder, and IgG subclass  
CC selection disorder. The DNA sequences encoding AID may be used for gene  
CC therapy and the antibodies to the AID protein may be used for diagnosis  
CC and treatment of these disorders  
XX  
SQ Sequence 2818 BP; 868 A; 548 C; 626 G; 776 T; 0 U; 0 Other;  
Query Match 100.0%; Score 271; DB 3; Length 2818;  
Best Local Similarity 100.0%; Pred. No. 4,5e-58;  
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AACGGCTGCCAGCTGGAAATGCTCTTCCTCCGCTACATCTGGAGCTGGAGCTAGACCTT 60  
DB 236 AACGGCTGCCAGCTGGAAATGCTCTTCCTCCGCTACATCTGGAGCTGGAGCTAGACCTT 295  
QY 61 GGGCGCTGTACCGCGCTACCTGTTCACTCTGTGAGCCCTGTACACTGTGGCCGA 120  
DB 296 GGGCGCTGTACCGCGCTACCTGTTCACTCTGTGAGCCCTGTACACTGTGGCCGA 355  
QY 121 CATGTGGCCGACTTTCTGGAGGGAACCCCACTGATGTAGATCTTACCGCGGC 180  
DB 356 CATGTGGCCGACTTTCTGGAGGGAACCCCACTGATGTAGATCTTACCGCGGC 415  
QY 181 CTCTACTCTGTGAGGAGCGGAGGCTGAGCCGAGGCGCTCGGCGGTGAGACCGGCC 240  
DB 416 CTCTACTCTGTGAGGAGCGGAGGCTGAGCCGAGGCGCTCGGCGGTGAGACCGGCC 475  
QY 241 GGGGTGCAATATAGCCATCATGACCTTCAAG 271  
DB 476 GGGGTGCAATATAGCCATCATGACCTTCAAG 506  
RESULT 7  
AAC55314  
ID AAC55314 standard; DNA; 6564 BP.  
XX  
AC AAC55314;  
XX  
DT 05-FEB-2001 (first entry)  
XX  
DE Human activation-induced cytidine deaminase genomic DNA SEQ ID NO:10.  
XX  
KW Activation-induced cytidine deaminase; AID; cytidine deaminase;  
KW immune related disease; allergy; allergic disease; antileukemic;  
KW antineoplastic; antineoplastic; ophthalmological; anti-HIV; dermatological;  
KW gene therapy; B cell associated immune system disorder; food allergy;  
KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
KW IgA nephritis; gamma-globulinemia; atopic dermatitis; allergic colitis;  
KW drug allergy; allergic rhinitis; Rosen disease; DiGeorge disease; AIDS;  
KW ataxia telangiectasia; common variable immunodeficiency disorder;  
KW major histocompatibility class II deficiency disease;  
KW auto immunodeficiency syndrome; IgG subclass selection disorder; ds.  
XX  
OS Homo sapiens.  
XX  
XX  
XX WO200058480-A1.  
XX  
XX PD 05-OCT-2000.  
XX  
XX PF 28-MAR-2000; 2000WO-JP001918.  
XX  
XX PR 29-MAR-1999; 99JP-00087192.  
XX PR 24-JUN-1999; 99JP-00178999.  
XX PR 27-DEC-1999; 99JP-00371382.  
XX  
XX (NIBB) JAPAN TOBACCO INC.  
XX (HONJ/) HONJO T.  
XX  
PI Honjo T, Muramatsu M;

XX  
DR WPI; 2000-611715/58.  
XX  
XX Nucleic acid encoding activation induced cytidine deaminase, useful as a  
PT target for drug development for immune-related diseases including  
PT allergies.  
XX  
PS Claim 17; Page 145-150; 174pp; Japanese.  
XX  
XX The present invention describes an activation-induced cytidine deaminase  
XX (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and has  
XX cytidine activity similar to APOBEC-1. AID has antileukemic, antineoplastic,  
XX antineoplastic, ophthalmological, anti-HIV and dermatological activities,  
XX and can be used in gene therapy. AID polynucleotides are useful in  
XX methods for identifying drugs for the treatment of B cell associated  
XX immune system disorders, immunodeficiency diseases and allergies, such as  
XX immunoglobulin A (IgA) deficiency disease, IgA nephritis, gamma-  
XX globulinemia, atopic dermatitis, allergic colitis, asthma, food allergy,  
XX drug allergy, allergic rhinitis, Rosen disease, DiGeorge disease, ataxia  
XX telangiectasia, common variable immunodeficiency disorder, MHC (major  
XX histocompatibility class II deficiency disease, AIDS (auto  
XX immunodeficiency syndrome), elevated IgE disorder, and IgG subclass  
XX selection disorder. The DNA sequences encoding AID may be used for gene  
XX therapy and the antibodies to the AID protein may be used for diagnosis  
XX and treatment of these disorders. The present sequence represents a  
XX genomic DNA sequence of human AID  
SQ Sequence 6564 BP; 1909 A; 1358 C; 1383 G; 1914 T; 0 U; 0 Other;  
Query Match 100.0%; Score 271; DB 3; Length 6564;  
Best Local Similarity 100.0%; Pred. No. 5.1e-58;  
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AACGGCTGCCAGCTGGAAATGCTCTTCCTCCGCTACATCTGGAGCTGGAGCTAGACCTT 60  
DB 2592 AACGGCTGCCAGCTGGAAATGCTCTTCCTCCGCTACATCTGGAGCTGGAGCTAGACCTT 2651  
QY 61 GGGCGCTGTACCGCGCTACCTGTTCACTCTGTGAGCCCTGTACACTGTGGCCGA 120  
DB 2652 GGGCGCTGTACCGCGCTACCTGTTCACTCTGTGAGCCCTGTACACTGTGGCCGA 2711  
QY 121 CATGTGGCCGACTTTCTGGAGGGAACCCCACTGATGTAGATCTTACCGCGGC 180  
DB 2712 CATGTGGCCGACTTTCTGGAGGGAACCCCACTGATGTAGATCTTACCGCGGC 2771  
QY 181 CTCTACTCTGTGAGGAGCGGAGGCTGAGCCGAGGCGCTCGGCGGTGAGACCGGCC 240  
DB 2772 CTCTACTCTGTGAGGAGCGGAGGCTGAGCCGAGGCGCTCGGCGGTGAGACCGGCC 2831  
QY 241 GGGGTGCAATATAGCCATCATGACCTTCAAG 271  
DB 2832 GGGGTGCAATATAGCCATCATGACCTTCAAG 2862  
RESULT 8  
AAC55339  
ID AAC55339 standard; DNA; 11204 BP.  
XX  
XX AAC55339;  
XX  
XX DT 05-FEB-2001 (first entry)  
XX  
XX DE Human activation-induced cytidine deaminase genomic DNA SEQ ID NO:35.  
XX  
XX KW Activation-induced cytidine deaminase; AID; cytidine deaminase;  
KW immune related disease; allergy; allergic disease; antileukemic;  
KW antineoplastic; antineoplastic; ophthalmological; anti-HIV; dermatological;  
KW gene therapy; B cell associated immune system disorder; food allergy;  
KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
KW IgA nephritis; gamma-globulinemia; atopic dermatitis; allergic colitis;  
KW drug allergy; allergic rhinitis; Rosen disease; DiGeorge disease; AIDS;  
KW ataxia telangiectasia; common variable immunodeficiency disorder;  
KW major histocompatibility class II deficiency disease;



KM auto immunodeficiency syndrome; IgG subclass selection disorder; ds.  
XX Homo sapiens.  
XX MO200058480-A1.  
XX  
XX 05-OCT-2000.  
XX  
XX 28-MAR-2000; 2000MO-JP001918.  
XX  
XX 29-MAR-1999; 99DP-00087192.  
XX 24-JUN-1999; 99DP-00178999.  
XX 27-DEC-1999; 99DP-00371382.  
XX  
XX (NISH ) JAPAN TOBACCO INC.  
XX (HONJ/) HONJO T.  
XX  
XX Honjo T, Muramatsu M;  
XX WPI; 2000-611715/58.  
XX  
XX Nucleic acid encoding activation induced cytidine deaminase, useful as a  
XX target for drug development for immune-related diseases including  
XX allergies.  
XX  
XX Claim 17; Page 163-170; 174pp; Japanese.  
XX  
XX The present invention describes an activation-induced cytidine deaminase  
XX (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and has  
XX cytidine activity similar to APOBEC-1. AID has anti-allergic, anti-nausea,  
XX antispasmodic, ophthalmological, anti-HIV and dermatological activities,  
XX and can be used in gene therapy. AID polynucleotides are useful in  
XX methods for identifying drugs for the treatment of B cell associated  
XX immune system disorders, immunodeficiency diseases and allergies, such as  
XX immunoglobulin A (IGA) deficiency disease, IGA nephritis, gamma-  
XX globulinemia, atopic dermatitis, allergic colitis, asthma, food allergy,  
XX drug allergy, allergic rhinitis, allergic disease, Digeorge disease, ataxia  
XX telangiectasia, common variable immunodeficiency disorder, MHC (major  
XX histocompatibility class II deficiency disease, AIDS (auto  
XX immunodeficiency syndrome), elevated Ige disorder, and IgG subclass  
XX selection disorder. The DNA sequences encoding AID may be used for gene  
XX therapy and the antibodies to the AID protein may be used for diagnosis  
XX and treatment of these disorders. The present sequence represents a  
XX genomic DNA sequence of human AID  
XX  
XX Sequence 11204 BP; 3305 A; 2273 C; 2373 G; 3253 T; 0 U; 0 Other;

Query Match 100.0%; Score 271; DB 3; Length 11204;  
Best Local Similarity 100.0%; Pred. No. 5,5e-58;  
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AACGGCTGACAGTGAATGCTCTCTCCGCTACATCTCGAGCTGAGACTAGACCT 60  
DB 7807 AACGGCTGACAGTGAATGCTCTCTCCGCTACATCTCGAGCTGAGACTAGACCT 7866  
QY 61 GGGCGGTGCTACCGCGCTCACTGTTCACTCTCTGAGGCCCTGCTACAGACTGTGCCGA 120  
DB 7867 GGGCGGTGCTACCGCGCTCACTGTTCACTCTCTGAGGCCCTGCTACAGACTGTGCCGA 7926  
QY 121 CATGTGGCGCACTTTCTGCGAGGAAACCCCACTTACGTCTGAGATCTTCAACGGCGGC 180  
DB 7927 CATGTGGCGCACTTTCTGCGAGGAAACCCCACTTACGTCTGAGATCTTCAACGGCGGC 7986  
QY 181 CTCTACTCTGTGAGAGCGGAAAGCTGAGCCCGGAGGGGCTGGGCGCTGACCGCGGC 240  
DB 7987 CTCTACTCTGTGAGAGCGGAAAGCTGAGCCCGGAGGGGCTGGGCGCTGACCGCGGC 8046  
QY 241 GGGGTCAATATAGCATCATGACCTTCAAG 271  
DB 8047 GGGGTCAATATAGCATCATGACCTTCAAG 8077

RESULT 9

AB573286  
ID AB573286 standard; DNA; 11204 BP.  
XX  
XX AC AB573286;  
XX  
XX 04-DEC-2002 (first entry)  
XX  
XX DNA encoding human translocation del(12p) protein #1.  
XX  
XX  
XX Chromosome aberration; oncogenic fusion protein; cancer;  
XX proliferative disease; cellular protein isoform; heat shock protein 90;  
XX HSP-90; rheumatoid arthritis; cancer; haematopoietic disorder;  
XX T cell lymphoma; B cell lymphoma; chronic myeloid leukaemia; CML;  
XX acute myeloid leukaemia; ALL; chronic myelomonocytic leukaemia; CMML;  
XX acute lymphoblastic leukaemia; ALL; APL; NHL; solid tumour;  
XX papillary thyroid carcinoma; Ewing's sarcoma; melanoma; liposarcoma;  
XX rhabdomyosarcoma; synovial sarcoma; viral infection; gene; ds.  
XX  
XX Homo sapiens.  
XX  
XX MO200269900-A2.  
XX  
XX 12-SEP-2002.  
XX  
XX 01-MAR-2002; 2002MO-US006518.  
XX  
XX 01-MAR-2001; 2001US-0272751P.  
XX  
XX (CONF-) CONFORMA THERAPEUTICS CORP.  
XX  
XX Fittz LC, Burrows FU;  
XX WPI; 2002-698710/75.  
XX P-PSDB; ABG95082.  
XX  
XX Treating genetically-defined disease associated with chromosomal  
XX aberrations yielding oncogenic fusion proteins, e.g. cell proliferative  
XX diseases, involves administering an inhibitor of heat shock protein 90.  
XX  
XX Disclosure; Page 242-245; 389pp; English.  
XX  
XX The invention describes a method of treating genetically-defined disease  
XX associated with chromosomal aberrations yielding oncogenic fusion  
XX proteins (I), treating cancerous cells containing (I) in a heterogeneous  
XX cell population, treating proliferative diseases associated with mutant  
XX protein or cellular protein isoforms (II) dependent on heat shock protein  
XX (HSP)-90, or selectively treating cells expressing (II) involving  
XX administering HSP90-inhibitor. The method is useful for treating  
XX genetically-defined disease with chromosomal aberration yielding  
XX oncogenic fusion protein, treating cancerous cells containing fusion  
XX protein in heterogeneous cell population, treating proliferative disease  
XX (e.g. rheumatoid arthritis or cancer) associated with mutant protein or  
XX cellular protein isoform dependent on heat shock protein (HSP)-90 (e.g.  
XX p53), or selectively treating cells expressing mutant protein or cellular  
XX protein isoform in a patient heterozygous for (II). The method is useful  
XX for treating a disease e.g. haematopoietic disorder such as T or B cell  
XX lymphoma, chronic myeloid leukaemia (CML), APL, ALL, NHL and CMML,  
XX or a disease characterised by a solid tumour such as papillary thyroid  
XX carcinoma, Ewing's sarcoma, melanoma, liposarcoma, rhabdomyosarcoma and  
XX synovial sarcoma. The method is also useful for treating viral  
XX infections. This represents the DNA sequence of a chromosome aberration  
XX  
XX Sequence 11204 BP; 3305 A; 2273 C; 2373 G; 3253 T; 0 U; 0 Other;

Query Match 100.0%; Score 271; DB 6; Length 11204;  
Best Local Similarity 100.0%; Pred. No. 5,5e-58;  
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AACGGCTGACAGTGAATGCTCTCTCCGCTACATCTCGAGCTGAGACTAGACCT 60  
DB 7807 AACGGCTGACAGTGAATGCTCTCTCCGCTACATCTCGAGCTGAGACTAGACCT 7866  
QY 61 GGGCGGTGCTACCGCGCTCACTGTTCACTCTCTGAGGCCCTGCTACAGACTGTGCCGA 120



```

Db      7867 GGGCCCTGCTACCGCGTCACTGTTCACTCTGAGACCCCTGTACAGACTGTGCCGA 7926
QY      121 CATGTGGCCGACATTTCTGTGGAGGAAACCCCACTCAGTGTGAGATTTTACCGCGCGC 180
Db      7927 CATGTGGCCGACATTTCTGTGGAGGAAACCCCACTCAGTGTGAGATTTTACCGCGCGC 7986
QY      181 CTCTACTTCTGTGAGGACCGCAAGCTGAGCCCGAGGCGCTGCGCGCTCACCGCGCC 240
Db      7987 CTCTACTTCTGTGAGGACCGCAAGCTGAGCCCGAGGCGCTGCGCGCTCACCGCGCC 8046
QY      241 GGGGTGCATATAGCCATCATGACCTTCAAG 271
Db      8047 GGGGTGCATATAGCCATCATGACCTTCAAG 8077

RESULT 10
AAC55307
ID      AAC55307 standard; cDNA; 2440 BP.
XX
AC      AAC55307;
XX
DT      05-FEB-2001 (first entry)
XX
DE      Mouse activation-induced cytidine deaminase encoding cDNA SEQ ID NO:1.
XX
XX      Activation-induced cytidine deaminase; AID; cytidine deaminase;
XX      immune related disease; allergy; allergic disease; antiallergic;
XX      antianemia; antiasthmatic; ophthalmological; anti-HIV; dermatological;
XX      gene therapy; B cell associated immune system disorder; food allergy;
XX      immunodeficiency disease; immunoglobulin A deficiency disease; asthma;
XX      IgA nephritis; gamma-globulinemia; atopic dermatitis; allergic colitis;
XX      drug allergy; allergic rhinitis; Rosen disease; Disgeorge disease; AIDS;
XX      ataxia telangiectasia; common variable immunodeficiency disorder;
XX      major histocompatibility class II deficiency disease;
XX      auto immunodeficiency syndrome; IgG subclass selection disorder; ss.
XX
OS      Mus musculus.
XX
FH      Key      Location/Qualifiers
FT      CDS      93..689
FT      FT      /tag= a
FT      FT      /product= "activation-induced cytidine deaminase"
XX
PN      WO200058480-A1.
XX
PD      05-OCT-2000.
XX
PF      28-MAR-2000; 2000WO-JP001918.
XX
XX      29-MAR-1999; 99JP-00087192.
XX      24-JUN-1999; 99JP-00178999.
XX      27-DEC-1999; 99JP-00371382.
XX
PA      (NISR ) JAPAN TOBACCO INC.
PA      (HONJ/) HONJO T.
XX
PI      Honjo T, Muramatsu M;
XX
DR      WPI, 2000-611715/58.
XX      P-PSDB; AAB24197.
XX
XX      Nucleic acid encoding activation induced cytidine deaminase, useful as a
XX      target for drug development for immune-related diseases including
XX      allergies.
XX
PS      Claim 3; Page 126-130; 174pp; Japanese.
XX
XX      The present sequence encodes mouse activation-induced cytidine deaminase
XX      (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and has
XX      catalytic activity similar to APOBEC-1. AID has antiallergic, antianemic,
XX      antiasthmatic, ophthalmological, anti-HIV and dermatological activities,
XX      and can be used in gene therapy. AID polynucleotides are useful in

```

```

CC      methods for identifying drugs for the treatment of B cell associated
CC      immune system disorders, immunodeficiency diseases and allergies, such as
CC      immunoglobulin A (IGA) deficiency disease, IgA nephritis, gamma-
CC      globulinemia, atopic dermatitis, allergic colitis, asthma, food allergy,
CC      drug allergy, allergic rhinitis, rosen disease, Disgeorge disease, ataxia
CC      telangiectasia, common variable immunodeficiency disorder, MHC (major
CC      histocompatibility class) class II deficiency disease, AIDS (auto
CC      immunodeficiency syndrome), elevated IgE disorder, and IgG subclass
CC      selection disorder. The DNA sequences encoding AID may be used for gene
CC      therapy and the antibodies to the AID protein may be used for diagnosis
CC      and treatment of these disorders
XX
SQ      Sequence 2440 BP; 706 A; 546 C; 551 G; 636 T; 0 U; 1 Other;
XX
Query Match      81.8%; Score 221.6; DB 3; Length 2440;
Best Local Similarity 89.2%; Pred. No. 9.2e-46;
Matches 239; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
QY      4 GGGTGCACGTGATGCTCTCTCCCTACATCTCGAGTGGAGACCTAGACCTTGC 63
Db      252 GGGTGCACGTGATGCTCTCTCCCTACATCTCGAGTGGAGACCTAGACCTTGC 311
QY      64 CGCTGTACCGGTCACTGTGTTACCTCTGAGGCCCTGTGACACTGTGCCGACAT 123
Db      312 CGGTGTACCGGTCACTGTGTTACCTCTGAGGCCCTGTGACACTGTGCCGACAT 371
QY      124 GTGGCCGACCTTTCTGTGGAGGAAACCCCACTCAGTGTGAGATCTTACCGCGCGCTC 183
Db      372 GTGGCTGAGTTTCTGTGAGGAAACCCCACTCAGTGTGAGATTTTACCGCGCGCTC 431
QY      184 TACTTCTGTGAGGACCGCAAGCTGAGCCCGAGGCGCTCGCGCTGACCGCGCGCTC 243
Db      432 TACTTCTGTGAGGACCGCAAGCTGAGCCCGAGGCGCTCGCGCTGACCGCGCGCTC 491
QY      244 GTGCAATATAGCCATCATGACCTTCAAG 271
Db      492 GTGCAATATAGCCATCATGACCTTCAAG 519

RESULT 11
ABK83568
ID      ABK83568 standard; DNA; 201143 BP.
XX
AC      ABK83568;
XX
DT      29-AUG-2002 (first entry)
XX
DE      Human DNA differentially expressed in granulocytic cells #139.
XX
XX      Human; de; granulocytic cell; DNA chip; bacterial infection;
XX      viral infection; parasitic infection; protozoal infection;
XX      fungal infection; sterile inflammatory disease; psoriasis;
XX      rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
XX      cardiac reperfusion injury; renal reperfusion injury; ARDS;
XX      adult respiratory distress syndrome; inflammatory bowel disease;
XX      Crohn's disease; ulcerative colitis; periodontal disease;
XX      granulocyte activation; chronic inflammation; allergy.
XX
OS      Homo sapiens.
XX
XX      WO200228999-A2.
XX
XX      11-APR-2002.
XX
XX      03-OCT-2001; 2001WO-US030821.
XX
XX      03-OCT-2000; 2000US-0237189P.
XX
XX      (GENE-) GENE LOGIC INC.
XX
XX      Beazer-Barclay Y, Weisman SM, Yamaga S, Vockley J;
XX
XX      WPI; 2002-435328/46.

```



XX Detecting granulocyte activation by detecting differential expression of  
PT genes associated with granulocyte activation, which serves as diagnostic  
PT markers that is useful for monitoring disease states and drug toxicity.

PS Claim 1; SEQ ID NO 139; 114bp; English.

XX The invention relates to detecting (M1) granulocyte (GC) activation  
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by  
CC DNA chip analysis as given in the specification, and comparing the  
CC expression level to an expression level in an unactivated GC, where  
CC differential expression of Gs is indicative of GCA. Also included are  
CC modulating (M2) GCA by contacting GC with an agent that alters the  
CC expression of at least one gene in Gs; (2) screening (M3) for an agent  
CC capable of modulating GCA or an inflammation (especially chronic) in a  
CC tissue, an allergic response in a subject, exposure of a subject to a  
CC pathogen or sterile inflammatory disease using the gene expression  
CC profile; (3) detecting (M4) an inflammation (especially chronic) in a  
CC tissue, an allergic response in a subject, exposure of a subject to a  
CC pathogen or sterile inflammatory disease, by detecting the level of  
CC expression in a sample of the tissue of gene(s) from Gs, where the level  
CC of expression of the gene is indicative of inflammation; (4) treating  
CC (M5) an inflammation (especially chronic) or in a tissue, an allergic  
CC response in a subject, exposure of a subject to a pathogen or sterile  
CC inflammatory disease, by contacting a tissue having inflammation with an  
CC agent that modulates the expression of gene(s) from Gs in the tissue. M1  
CC is useful for detecting GCA; M2 is useful for modulating GCA; M3 is useful  
CC for screening an agent capable of modulating GCA preferably in an  
CC inflammation in a tissue; M4 is useful for detecting an inflammation  
CC (especially chronic) in a tissue, an allergic response in a subject,  
CC exposure of a subject to a pathogen or sterile inflammatory disease (e.g.  
CC psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis,  
CC cardiac reperfusion injury, renal reperfusion injury, ARDS, adult  
CC respiratory distress syndrome, inflammatory bowel disease, Crohn's  
CC disease, ulcerative colitis, periodontal disease; also bacterial  
CC infection, viral infection, parasitic infection, protozoal infection,  
CC fungal infection and M5 is useful for treating one of the above  
CC conditions. The present sequence represents a gene differentially  
CC expressed in granulocytes. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pat\_sequences

XX Sequence 201143 BP; 50606 A; 49308 C; 49683 G; 51546 T; 0 U; 0 Other;  
SQ

Query Match 32.3%; Score 87.4; DB 6; Length 201143;  
Best Local Similarity 60.8%; Pred. No. 4.9e-12;  
Matches 161; Conservative 0; Mismatches 101; Indels 3; Gaps 1;

XX 7 TGCCAGGTGGAATTGCTCTTCCCTCCGCTACATCTGGAGCTGGAGACCTTGAACCTGGCCG 66  
DB 167899 TGTCAATGCAAGAGGCTCTCTCTCTGTTGTGAGACATATCTCTCTTAACACA 167858

XX 67 TGCTACCGGCTCACTGTTCACTCTCTGAGGCCCTGTACGACATGTGCCGACATGTG 126

DB 167959 AACTTAGAGAGTCACTGTTCACTCTGAGGCCCTGTCCGACAGTGTCCAGGAGGTG 168018

XX 127 GCCGACTTTCTGCGAGGAAACCCCAACTCAGTCTGAGAGTTCTTACCGCGGCTCTTAC 186

DB 168019 GCCGAGTTCTTGCGAGGACCAACAGTAATCTCACCATCTTACCGCCGCTCTAC 168078

XX 187 TTCTGTGAGGACCGCAAGGTGAGCCCGAGGGCTGCGCGGTGTGACCGCGCCGGGGTG 246

DB 168079 TACTTCTGGGAA--TTCAGATTACCGAGGAGGCTCCGAGCTGAGTGAAGAAAGGGGCTC 168135

XX 247 CAATAGCCATGACCTTCAAG 271

DB 168136 TCCGTGAGATCATGGCTACAAAG 168160

RESULT 12  
AAV48231  
ID AAV48231 standard; cDNA; 610 BP.

XX AAV48231;  
AC 09-NOV-1998 (first entry)  
DT  
XX  
XX  
DE Human RNA editing enzyme nucleotide sequence.

XX ss: human; RNA editing enzyme; REE; pharmaceutical carrier; cancer;  
KM viral diseases; circulatory system disorder; RNA processing; B.  
KM hypercholesterolemia; alpha-galactosidase; apolipoprotein B.

OS Homo sapiens.

XX Key Location/Qualifiers  
FT 1..573  
FT /tag= a  
FT /product= "RNA editing enzyme"

XX US5804185-A.

XX 08-SEP-1998.

XX 13-MAR-1997; 97US-00816241.

XX 13-MAR-1997; 97US-00816241.

XX (INCY-) INCYTE PHARM INC.

XX Bandman O, Gold SK,

XX WPI; 1998-505585/43.

XX P-PSDB; AAW77092.

XX Human RNA editing enzyme and poly(nucleotide(s) encoding it - useful for  
PT recombinant production of the enzyme and treatment and detection of  
PT disorders associated with incorrect RNA processing.

XX Claim 4; Fig 1; 27pp; English.

XX The human RNA editing enzyme (REE) is used in a pharmaceutical carrier  
CC for the treatment of cancer, viral diseases and circulatory system  
CC disorders. The enzyme is used in vivo for the correct processing of RNA  
CC transcripts of genes e.g. change of a codon in apolipoprotein B (apob)  
CC RNA to give a 100 and 48 kDa product transcribed from the same gene.  
CC Certain disorders have been linked to incorrect RNA editing, e.g. failure  
CC of apo B editing leads to excessive apob 100 production and  
CC hypercholesterolemia. Other disorders thought to be linked to incorrect  
CC RNA processing include aberrant alpha-galactosidase processing in Fabry's  
CC disease and neurofibromatosis type I. The enzyme can be produced  
CC recombinantly to treat related disorders. It can also be used to raise  
CC antibodies for immuno-based detection of REE expression levels e.g. ELISA

SQ Sequence 610 BP; 158 A; 154 C; 154 G; 144 T; 0 U; 0 Other;

Query Match 31.7%; Score 85.8; DB 2; Length 610;  
Best Local Similarity 60.4%; Pred. No. 5.3e-12;  
Matches 160; Conservative 0; Mismatches 102; Indels 3; Gaps 1;

XX 7 TGCCAGGTGGAATTGCTCTTCCCTCCGCTACATCTGGAGCTGGAGACCTTGAACCTGGCCG 66

DB 193 TGTCAATGCAAGAGGCTCTCTCTCTGTTGTGAGACACATATCTCTCTTAACACA 252

XX 67 TGCTACCGGCTCACTGTTCACTCTCTGAGGCCCTGTACGACATGTGCCGACATGTG 126

DB 253 AAGTACAGGCTCACTGTTCACTCTCTGAGGCCCTGTCCGACAGTGTGAGAGGAGTG 312

XX 127 GCCGACTTTCTGCGAGGAAACCCCAACTCAGTCTGAGAGTTCTTACCGCGGCTCTTAC 186

DB 313 GCCGAGTTCTTGCGAGGACCAACAGTAATCTCACCATCTTACCGCCGCTCTTAC 372

XX 187 TTCTGTGAGGACCGCAAGGTGAGCCCGAGGGCTGCGCGGTGTGACCGCGCCGGGGTG 246

DB 373 TACTTCCAGTATC--CATGTTACAGGAGGGGCTCCGAGCTGAGTGAAGAAAGGGGTC 429



QY 247 CAATAGCCATCATGACCTTCAAAG 271  
DB 430 GCTGTGAGATCATGACTATGAG 454

RESULT 13  
ID AAA72058  
AAA72058 standard; cDNA; 610 BP.

AC AAA72058;  
DT 24-NOV-2000 (first entry)

DE cDNA encoding human RNA editing enzyme REE-2.

KW RNA editing enzyme; REE-2; human; HEPR homologue; REPR homologue;  
KW photobolilin I homologue; cancer; tumour; autoimmune disorder;  
KW circulatory system disorder; hypercholesterolemia; viral infection;  
KW neurological disease; neurofibromatosis; transcript editing; detection;  
ss.

OS Homo sapiens.  
XX  
XX  
FH Key Location/Qualifiers  
FT CDS 1..573  
FT /tag=a  
FT /product= "Human REE-2"

PN US6087108-A.  
XX  
PD 11-JUL-2000.  
XX  
PF 03-AUG-1998; 98US-00128395.  
XX  
PR 13-MAR-1997; 97US-00816241.  
XX  
PA (INCY-) INCYTE PHARM INC.  
XX  
PI Goli SK, Bandman O;  
XX  
DR MPI: 2000-531340/48.  
DR P-PSDB; AAB11973.  
XX  
PT Detecting polynucleotide encoding human RNA editing enzyme comprising  
PT hybridizing an isolated and purified polynucleotide complementary to the  
PT polynucleotide and detecting the hybridization complex.  
XX  
PS Claim 5; Fig 1A-B; 27bp; English.  
XX  
XX This sequence represents the cDNA encoding human RNA editing enzyme REE-  
CC 2. cDNA encoding REE-2 was initially isolated in a prostate tumour cDNA  
CC library, with the present sequence representing a consensus. REE-2 has  
CC chemical and structural homology with the human apob mRNA editing protein  
CC HEPR (28% identity), the rat HEPR homologue REPR (30% identity), and a  
CC portion of the mRNA editing enzyme photobolilin I (43% identity). REE-2 was  
CC found to be expressed in a variety of cDNA libraries, a high proportion  
CC of which were derived from tumours, neuronal tissues, immune system cells  
CC or synovial tissue from arthritis patients. REE-2 is therefore thought to  
CC be associated with the development of cancer, autoimmune disorders,  
CC circulatory system disorders (e.g. hypercholesterolemia), viral  
CC infections and neurological diseases (e.g., neurofibromatosis). REE-2 or  
CC its nucleic acids may be used in the diagnosis, treatment and prevention  
CC of such diseases via the modulation of transcript editing, which in turn  
CC has effects on the encoded protein (e.g., an alteration in protein  
CC activity). The invention specifically relates to methods of detecting  
CC nucleic acids encoding human REE-2 in a biological sample  
XX  
XX Sequence 610 BP; 158 A; 154 C; 154 G; 144 T; 0 U; 0 Other;

Query Match 31.7%; Score 85.8; DB 3; Length 610;  
Best Local Similarity 60.4%; Pred. No. 5,3e-12;  
Matches 160; Conservative 0; Mismatches 102; Indels 3; Gaps 1;

QY 7 TGGCAGTGAATTTGCTCTTCCGCTACATGAGCTGGAGACCTAGACCTGGCCG 66  
DB 193 TGTATATGAGAAAGTGTCTTCTTCTTGGTTTGAGAGACATGACTGTCTTAACA 252

QY 67 TGCATCCGCTGACCTGTTCACTCTGAGAGCCCTGCTACGACTGTGCCGACATGTG 126  
DB 253 AAGTACAGGTCACTGTGATCATATCTGGAGCCCTTGGCCAGACTGTCCAGGAGGTG 312

QY 127 GCCGACTTTTGGGAGGAAACCCAACTCAGTCTGAGATTTTCAACCGGCGCTTAC 186  
DB 313 GCCGAGTTCCTGGCCAGGACAGCAAGTGAATCTCACCATCTTCAACCGCGCTTAC 372

QY 187 TTCTGTGAGAGACCGCAAGGCTGAGCCCGAGGGGCTGGCGGTGACCCGCGGGGTG 246  
DB 373 TACTTCCAGTATC--CATGTTACAGAGAGGGGCTCCGAGGCTGAGTACAGAGAGGGGTG 429

QY 247 CAATAGCCATCATGACCTTCAAAG 271  
DB 430 GCTGTGAGATCATGACTATGAG 454

RESULT 14  
ID AAS41669  
AAS41669 standard; cDNA; 819 BP.

AC AAS41669;  
DT 17-DEC-2001 (first entry)

DE cDNA encoding novel human enzyme polypeptide #885.  
XX  
XX  
XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;  
KW ligase; hyperproliferative disorder; immunodeficiency disorder;  
KW autoimmune disorder; neurological disorder; metabolic disorder;  
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;  
KW blood-related disorder; infectious disorder; gene therapy; cytostatic;  
KW anti arthritic; nephrotropic; antioagulant; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200155301-A2.  
XX  
XX 02-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US001239.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
XX 04-FEB-2000; 2000US-0180628P.  
XX 24-FEB-2000; 2000US-0184664P.  
XX 02-MAR-2000; 2000US-0186350P.  
XX 16-MAR-2000; 2000US-0189874P.  
XX 17-MAR-2000; 2000US-0190076P.  
XX 18-APR-2000; 2000US-0198123P.  
XX 19-MAY-2000; 2000US-0205515P.  
XX 07-JUN-2000; 2000US-0209467P.  
XX 28-JUN-2000; 2000US-0214866P.  
XX 30-JUN-2000; 2000US-0215135P.  
XX 07-JUL-2000; 2000US-0216647P.  
XX 07-JUL-2000; 2000US-0216880P.  
XX 11-JUL-2000; 2000US-0217487P.  
XX 11-JUL-2000; 2000US-0217496P.  
XX 14-JUL-2000; 2000US-0218290P.  
XX 26-JUL-2000; 2000US-0220963P.  
XX 26-JUL-2000; 2000US-0220964P.  
XX 14-AUG-2000; 2000US-0224518P.  
XX 14-AUG-2000; 2000US-0224519P.  
XX 14-AUG-2000; 2000US-0225213P.  
XX 14-AUG-2000; 2000US-0225214P.  
XX 14-AUG-2000; 2000US-0225266P.  
XX 14-AUG-2000; 2000US-0225267P.  
XX 14-AUG-2000; 2000US-0225268P.  
XX 14-AUG-2000; 2000US-0225270P.







Best Local Similarity 60.4%; Pred. No. 5,6e-12;  
Matches 160; Conservative 0; Mismatches 102; Indels 3; Gaps 1;

Qy	7	TGCGACGTGAATTGCTCTTCTCCGCGTCACTCGAGACTGGAACTTGACCTTGCGCG	66
Db	291	TGTGATGCGAAGAAGTGCTTCCCTCTTGTGTTCTGCGACGACATGCTGTCTCTTAACA	350
Qy	67	TGCTACCGCGCTCACTGGTTCACCTTCCTGAGCCCTGTGACGACTGTGCCGACATGTG	126
Db	351	AAGTACCGAGGTCTACCTGTGTAACATTTTGAACTCTTGCCAGACTGTGCGAGGAAAGTG	410
Qy	127	GCGACTTTGTGAGGAGAAACCCCAACTCAGTCTGAGGATTTTCAACCGCGCGCTTAC	186
Db	411	GCGAGTTCCTGGCGAGGACAGGAACTGAATCTCAACATTTTACCGCCCGGCTTAC	470
Qy	187	TTCTGTGAGGACCGCAAGCTGAGCCCGAGGGGCTGCGCGGCTGACCGCGCCGGGGTG	246
Db	471	TACTTCCAGTATC---CATGTTACCGAGAGGGGCTCCGACGCTGAGTCAGAAAGGGGCTC	527
Qy	247	CAAAATAGCATCATGACCTTCAAG	271
Db	528	GCTGTGAGATCAATGACATGTGAAG	552

RESULT 15

ID ABL99876 standard; cDNA; 944 BP.

AC ABL99876;

DT 03-OCT-2002 (first entry)

Human secretory polynucleotide (sptm) 131

KM Human; s; gene; secretory protein; secretory polynucleotides; SPM;  
KM SPM-related disease; somatic gene therapy; germ-line gene therapy;  
KM severe combined immunodeficiency; intracellular parasite protection;  
KM fungal parasite; protozoan parasite; cell proliferative disorder; cancer;  
KM immune disorder; AIDS; neurological disorder; Parkinson's disease;  
KM motor neuron disorder; demyelinating disease; multiple sclerosis;  
KM meninges; abscess; prion disease; cerebral palsy;  
KM neuroskeletal disorder; peripheral nervous system disorder;  
KM dermatomyositis; polymyositis; myopathy; myasthenia gravis;  
KM mental disorder; Tourette's syndrome.

OS Homo sapiens.

PN WO200220756-A2.

PD 14-MAR-2002

30-AUG-2001; 2001WO-US027297.

PR 05-SEP-2000; 2000US-0229747P.  
PR 05-SEP-2000; 2000US-0229748P.  
PR 05-SEP-2000; 2000US-0229749P.  
PR 05-SEP-2000; 2000US-0229750P.  
PR 05-SEP-2000; 2000US-0229751P.  
PR 05-SEP-2000; 2000US-0230016P.  
PR 05-SEP-2000; 2000US-0230563P.  
PR 06-SEP-2000; 2000US-0230505P.  
PR 06-SEP-2000; 2000US-0230514P.  
PR 06-SEP-2000; 2000US-0230515P.  
PR 06-SEP-2000; 2000US-0230517P.  
PR 06-SEP-2000; 2000US-0230518P.  
PR 06-SEP-2000; 2000US-0230519P.  
PR 06-SEP-2000; 2000US-0230556P.  
PR 06-SEP-2000; 2000US-0230565P.  
PR 06-SEP-2000; 2000US-0230579P.  
PR 06-SEP-2000; 2000US-0230599P.  
PR 06-SEP-2000; 2000US-0230610P.  
PR 06-SEP-2000; 2000US-0230640P.  
PR 06-SEP-2000; 2000US-0230655P.

PR 06-SEP-2000; 2000US-0230988P.  
PR 06-SEP-2000; 2000US-0230989P.  
PR 06-SEP-2000; 2000US-0230990P.  
PR 07-SEP-2000; 2000US-0230896P.  
PR 07-SEP-2000; 2000US-0230897P.  
PR 07-SEP-2000; 2000US-0230951P.  
PR 07-SEP-2000; 2000US-0231163P.  
PR 07-SEP-2000; 2000US-0231832P.  
XX  
XA (INCY-) INCYTE GENOMICS INC.

PI Stuart J, Lincoln SE, Altus CM, Dufour GE, Chalup MS, Hillman JL;  
PI Jones AL, Yu JV, Wright RJ, Gjerdzen D, Liu TF, Yap PE, Dahl CR;  
PI Montiyama MG, Bradley DL, Rohagshi SD, Harris B, Roseberry AM;  
PI Gerstlin ER, Petalica CR, David MH, Panzer SR, Flores V, Datto A;  
PI Marwaha R, Chen AJ, Chang SC, Au AP, Inman RR;  
XX  
WPI: 2002-315658/35.  
DR P-PsDB: ABB97879.  
XX

PT Polynucleotide sequences encoding human secretory proteins useful for  
PT gene therapy of e.g. genetic deficiency disorders, cancers, and diseases  
PT caused by intracellular parasites.

PS Claim 1; Page 326; 585pp; English

CC The invention comprises the amino acid and coding sequences of human  
CC secretory (SPTM) proteins. The SPTM DNA and amino acid sequences are  
CC useful for treating a disease or condition associated with the expression  
CC of functional SPTM. The SPTM DNA sequences are useful for somatic or  
CC germ-line gene therapy to correct a genetic deficiency (e.g. severe  
CC combined immunodeficiency). The SPTM DNA sequences are also useful in  
CC providing protection against intracellular parasites (e.g. fungal  
CC parasites and protozoan parasites). The SPTM DNA and protein sequences  
CC are also useful for diagnosing cell proliferative disorders, cancer,  
CC immune disorders (e.g. AIDS), neurological disorders (e.g. Parkinson's  
CC disease), motor neuron disorders, demyelinating diseases (e.g. multiple  
CC sclerosis), meningitis, abscesses, prion diseases, cerebral palsy,  
CC neuroskeletal disorders, peripheral nervous system disorders,  
CC dermatomycosis and polymyositis, myopathy, myasthenia gravis, and mental  
CC disorders (e.g. Tourette's syndrome). cDNA sequences ABL939746 - ABL939929  
CC represent human secretory polynucleotides of the invention  
CC  
CC Sequence 944 BP; 222 A; 284 C; 219 G; 219 T; 0 U; 0 Other;  
CC  
CC

Sequence 944 BP; 222 A; 284 C; 219 G; 219 T; 0 U; 0 Other;

Query Match 31.7%; Score 85.8; DB 6; Length 944;

Matches 160; Conservative 0; Mismatches 102; Indels 3; Gaps 1;

QY	7	TGCCACGTGGAAATGGCTCTTCTCTCCGCTACATCTCGACCTGGACCTTGAACCTCTGGCCGC	66
Db	224	TGTATGACAGAAAGTGGCTTCTCTCTTGGTCTGGACGACATACGTGTCTCTTAACACA	283
QY	67	TGCTACCGCGCTCACTTGGTTCACTCTCTGGAGCCCCCTGCTACGACTGTGCCGACATGTG	126
Db	284	AAGTACCAAGGTCACCTGGTACACATCTTGGAGCCCTTGGCCGAGCTGTGCAAGGGGAGGTG	343
QY	127	GCCGACTTCTTGCAGGAGGAAACCCCACTCAGTGTGAGATCTTCAACCGGCGGCTCTAC	186
Db	344	GCCGAGTTCTGGSCACAGCACAGNAGTGAATTTCAACATCTTTCACCGGCCGCTCTAC	403
QY	187	TTCTGTGAGGACCGCAAGCTGAGCCCCGAGGGGCTGGCGGCGCTGCACCGCGCCGGGGTG	246
Db	404	TACTTCCAGTATC---CATGTTACCAAGAGGGGGCTCCGACGCTGAGTLCAGGAAGGGGTC	460
QY	247	CAATATGCATCATGACTTCAAG	271
Db	461	GCTGTGAGATCATGACTATGAAG	485

Search completed: March 12, 2004, 18:08:09  
Job time : 110.517 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: March 12, 2004, 17:39:25 ; Search time 20.1637 Seconds

(without alignments)  
7458.540 Million cell updates/sec

Title: US-09-966-880A-13

Perfect score: 271

Sequence: 1 aacggctgcacacgtgggaatt.....agccatcatgacattcaag 271

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 27747546 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:\*

- 1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*
- 2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*
- 3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq:\*
- 4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq:\*
- 5: /cgn2\_6/ptodata/2/ina/PCTUS.COMB.seq:\*
- 6: /cgn2\_6/ptodata/2/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	85.8	31.7	610	1	US-08-816-241-2
2	85.8	31.7	610	3	US-09-128-395-2
3	50.6	18.7	879	1	US-08-158-682A-3
4	44	16.2	891	1	US-08-687-895-2
5	44	16.2	891	2	US-09-040-483-2
6	41	15.1	879	1	US-08-158-682A-1
7	41	15.1	879	1	US-08-015-203-1
8	38	14.0	1929	4	US-09-380-420C-1
9	38	14.0	1929	4	US-09-899-642A-1
10	37.8	13.9	922	4	US-09-189-060B-73
11	36.4	13.4	624	4	US-09-252-991A-10053
12	36.4	13.4	1959	4	US-09-252-991A-9783
13	35.8	13.2	4689	3	US-09-105-537-34
14	35.8	13.2	36778	3	US-09-105-537-5
15	35.8	13.2	38506	3	US-09-320-878-19
16	35.8	13.2	38506	4	US-09-141-908-1
17	35.8	13.2	38506	4	US-09-657-440-19
18	35.4	13.1	2190	4	US-09-015-188-1
19	35.2	13.0	2584	4	US-09-716-129-47
20	34.6	12.8	1098	4	US-09-170-496D-117
21	34.6	12.8	1098	4	US-09-170-496D-225
22	34.6	12.8	1597	2	US-08-724-974A-1
23	34.6	12.8	1597	4	US-09-364-425B-26
24	34.4	12.7	2575	4	US-09-716-129-25
25	34.4	12.7	2888	4	US-09-016-434-1458
26	34.4	12.7	2980	4	US-09-266-225D-11
27	34	12.5	2505	4	US-09-334-818A-1

C 28	33.8	12.5	417	4	US-09-252-991A-9297	Sequence 9297, Ap
29	33.8	12.5	1215	4	US-09-252-991A-9259	Sequence 9259, Ap
C 30	33.8	12.5	1488	4	US-09-252-991A-9307	Sequence 9307, Ap
31	33.8	12.5	1602	4	US-09-252-991A-9270	Sequence 9270, Ap
32	33.8	12.5	15872	3	US-09-105-537-1	Sequence 1, Appli
33	33.8	12.5	15872	4	US-09-091-609-1	Sequence 1, Appli
34	33.8	12.5	15872	4	US-09-091-609-3	Sequence 3, Appli
C 35	33.8	12.5	4403765	3	US-09-103-840A-2	Sequence 2, Appli
C 36	33.8	12.5	4411529	3	US-09-103-840A-1	Sequence 1, Appli
37	33.6	12.4	1683	4	US-09-252-991A-8337	Sequence 8337, Ap
38	33.6	12.4	1833	4	US-09-252-991A-8371	Sequence 8371, Ap
C 39	33.6	12.4	1980	4	US-09-252-991A-8507	Sequence 8507, Ap
C 40	33	12.2	4911	4	US-09-718-852-1	Sequence 1, Appli
C 41	33	12.2	4911	4	US-09-718-852-1	Sequence 1, Appli
C 42	33	12.2	4911	4	US-09-718-852-1	Sequence 1, Appli
43	32.8	12.1	1017	4	US-09-252-991A-10831	Sequence 10831, A
44	32.6	12.0	1356	1	US-08-317-880-4	Sequence 4, Appli
45	32.6	12.0	1356	2	US-08-782-396-4	Sequence 4, Appli

## ALIGNMENTS

RESULT 1  
US-08-816-241-2  
Sequence 2, Application US/08916241  
Patent No. 5804185  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
TITLE OR INVENTION: NOVEL RNA EDITING ENZYME  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/816,241  
FILING DATE: Filed Herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0239 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 610 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: PROSTUT09  
CLONE: 1646833  
US-08-816-241-2  
Query Match 31.7%; Score 85.8; DB 1; Length 610;  
Best Local Similarity 60.4%; Pred. No. 1,6e-12;  
Matches 160; Conservative 0; Mismatches 102; Indels 3; Gaps 1;



QY 7 TGCACGAGAAATGCTCTTCCTCCGCTACATCTCGAGTGGAGCACTAGACCTCGGCGC 66  
DB 193 TGTATGACAGAAAGGTGCTTCTCTCTGTTTGTGCAACATATCTCTCTTAACA 252  
QY 67 TGTATCCGCGTCACTGTTGCTCACTCTGAGAGCCCTGCTCAAGACTGTGCGGACATGTG 126  
DB 253 AAGTACCGAGGTCACTGTTGACATCTTGAAGCCCTTGCCAGACTGTGAGGGAGGTG 312  
QY 127 GCCGACTTCTGCGAGGGAACCCCACTCACTGAGAGATCTTCAACCGCGGCTCTTAC 186  
DB 313 GCCGAGTCTCTGCGAGGGAACCCCACTCACTGAGAGATCTTCAACCGCGGCTCTTAC 312  
QY 187 TTTGTGAGAGCCGGAAGGCTGAGCCGAGGGGCTGAGGGGCTGACCGGCGGGGTG 246  
DB 373 TACTTCCAGTATC--CATGTTACAGAGAGGGGCTCCGACCTGTAGTCAAGAAAGGGGTC 429  
QY 247 CAATAGCCATCATGACCTTCAAG 271  
DB 430 GCTGTGAGATCATGACTATGAG 454

## RESULT 2

US-09-128-395-2  
Sequence 2, Application US/09128395  
Patent No. 6087108  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Goli, Surya K.  
TITLE OF INVENTION: NOVEL RNA EDITING ENZYME  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/128.395  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/816,241  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0239 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 610 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: PROSTITT09  
CLONE: 1646833  
US-09-128-395-2

Query Match 31.7%; Score 85.8; DB 3; Length 610;  
Best Local Similarity 60.4%; Pred. No. 1,6e-12;  
Matches 160; Conservative 0; Mismatches 102; Indels 3; Gaps 1;  
7 TSCCACTGGAATGCTCTTCTCCGCTACATCTCGAGTGGAGCTGAGACCTGAGCCCTGCGCC 66

DB 193 TGTATGACAGAAAGGTGCTTCTCTCTGTTTGTGCAACATATCTCTCTTAACA 252  
QY 67 TGTATCCGCGTCACTGTTGCTCACTCTGAGAGCCCTGCTCAAGACTGTGCGGACATGTG 126  
DB 253 AAGTACCGAGGTCACTGTTGACATCTTGAAGCCCTTGCCAGACTGTGAGGGAGGTG 312  
QY 127 GCCGACTTCTGCGAGGGAACCCCACTCACTGAGAGATCTTCAACCGCGGCTCTTAC 186  
DB 313 GCCGAGTCTCTGCGAGGGAACCCCACTCACTGAGAGATCTTCAACCGCGGCTCTTAC 312  
QY 187 TTTGTGAGAGCCGGAAGGCTGAGCCGAGGGGCTGAGGGGCTGACCGGCGGGGTG 246  
DB 373 TACTTCCAGTATC--CATGTTACAGAGAGGGGCTCCGACCTGTAGTCAAGAAAGGGGTC 429  
QY 247 CAATAGCCATCATGACCTTCAAG 271  
DB 430 GCTGTGAGATCATGACTATGAG 454

## RESULT 3

US-08-158-682A-3  
Sequence 3, Application US/08158682A  
Patent No. 5434058  
GENERAL INFORMATION:  
APPLICANT: Davidson, Nicholas O.  
TITLE OF INVENTION: Apolipoprotein B RNA Editing Protein:  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ARNOLD, WHITE & DURKEE  
STREET: 321 No. 5434058th Clark Street, Suite 800  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60610  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/158,682A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Cooley, Ronald B.  
REGISTRATION NUMBER: 27,187  
REFERENCE/DOCKET NUMBER: ARCD:085  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 744-0090  
TELEFAX: (312) 245-4961  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 879 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 23..730  
US-08-158-682A-3

Query Match 18.7%; Score 50.6; DB 1; Length 879;  
Best Local Similarity 59.3%; Pred. No. 0.00066;  
Matches 86; Conservative 0; Mismatches 59; Indels 0; Gaps 0;  
QY 40 TCGAGCTGGAGCTAGACCTGCGGCTGCTACCGGCTGACCTGTTCACTCTGAGAC 99  
DB 236 TCGAAGAGATTTTACCACTCATCACTGCTACACATCACTGCTGTGCTGAGGT 295  
QY 100 CCTGTACGACTGTGCGGAGCATGTGGCGGACTTTTTCGAGGGAACCCCACTCACT 159



Db 236 CCTGCTGGGATGCTCCAGGCTATTAGAGAGTTTGTGATCGGACCCCTGTGTACT 355  
QY 160 CTGAGATCTTACCGCGGCTCT 184  
Db 356 CTAGTATCTAGCTGCTGCTTT 380

## RESULT 4

US-08-687-895-2  
; Sequence 2, Application US/08687895  
; Patent No. 5747319

## GENERAL INFORMATION:

APPLICANT: Au-Young, Janice  
APPLICANT: Hawkins, Phillip R.  
APPLICANT: Hillman, Jennifer L.  
TITLE OF INVENTION: A NOVEL HUMAN MRNA EDITING ENZYME  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: U.S.  
ZIP: 94304

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/687,895

FILING DATE: Filed Herewith  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0109 US

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:  
LENGTH: 891 base pairs  
TYPE: nucleic acid

STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: cDNA  
IMMEDIATE SOURCE:  
LIBRARY: MUSCNOT1

CLONE: 57953  
US-08-687-895-2

Query Match 16.2%; Score 44; DB 1; Length 891;  
Best Local Similarity 53.6%; Pred. No. 0.027;  
Matches 89; Conservative 1; Mismatches 76; Indels 0; Gaps 0;

QY 49 GACCTAGACCTGCGCGCTGTAACCGCTACCTGCTTCACTCCGAGGCCCTGCTAC 108  
Db 480 GCGTTCGACCCAGCCCTGCGGTACATGTCACCTGTATGTCTCCAGCCCTGTGCA 539  
QY 109 GACTGTGCCGACATGTGCGCGCATTTCTGCGAGGGAACCCCACTCAGTGTGAGATC 168  
Db 540 GCGTGTCTGACCGCATTAACAACCTTAGCAAGACCAAGAACTGTGTCTCATTT 599  
QY 169 TTCACCGCGCGCTTACTTCTGTGTGAGACCGGAGGCTGAGCCCG 214  
Db 600 CTGCTGGTGTGACTTCTCATGTGTGAGAGACCGGAGATCCAGGCTG 645

RESULT 5  
US-09-040-482-2  
; Sequence 2, Application US/09040482  
; Patent No. 5916556  
; GENERAL INFORMATION:

APPLICANT: Au-Young, Janice  
APPLICANT: Hawkins, Phillip R.  
APPLICANT: Hillman, Jennifer L.  
TITLE OF INVENTION: A NOVEL HUMAN MRNA EDITING ENZYME  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: U.S.  
ZIP: 94304

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/040,482

FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/687,895

FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0109 US

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:  
LENGTH: 891 base pairs  
TYPE: nucleic acid

STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: cDNA  
IMMEDIATE SOURCE:  
LIBRARY: MUSCNOT1

CLONE: 57953  
US-09-040-482-2

Query Match 16.2%; Score 44; DB 2; Length 891;  
Best Local Similarity 53.6%; Pred. No. 0.027;  
Matches 89; Conservative 1; Mismatches 76; Indels 0; Gaps 0;

QY 49 GACCTAGACCTGCGCGCTGTAACCGCTACCTGCTTCACTCCGAGGCCCTGCTAC 108  
Db 480 GCGTTCGACCCAGCCCTGCGGTACATGTCACCTGTATGTCTCCAGCCCTGTGCA 539  
QY 109 GACTGTGCCGACATGTGCGCGCATTTCTGCGAGGGAACCCCACTCAGTGTGAGATC 168  
Db 540 GCGTGTCTGACCGCATTAACAACCTTAGCAAGACCAAGAACTGTGTCTCATTT 599  
QY 169 TTCACCGCGCGCTTACTTCTGTGTGAGACCGGAGGCTGAGCCCG 214  
Db 600 CTGCTGGTGTGACTTCTCATGTGTGAGAGACCGGAGATCCAGGCTG 645

## RESULT 6

US-08-158-682A-1  
; Sequence 1, Application US/08158682A  
; Patent No. 5434058

## GENERAL INFORMATION:

APPLICANT: Davidson, Nicholas O.  
TITLE OF INVENTION: Apolipoprotein B RNA Editing Protein:  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ARNOLD, WHITE & DURKEE  
STREET: 321 No. 5434058th Clark Street, Suite 800  
CITY: Chicago  
STATE: Illinois



COUNTRY: USA  
ZIP: 60610  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/158,682A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Colley, Ronald B.  
REGISTRATION NUMBER: 27,187  
REFERENCE/DOCKET NUMBER: ARCD:085  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 744-0090  
TELEFAX: (312) 245-4861  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 879 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-158-682A-1

Query Match  
Best Local Similarity 55.2%; Score 41; DB 1; Length 879;  
Matches 80; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

Db 41 CGAGCTGGAGCTTACCGCTGCGCGCTGCTACCGGCTGCTACCTGTTCACTCTGGAGCC 100  
255 CAGAAAGATACCTTTGTCTCAAAACACAGATGCTCATTAACCTGTTCTGTCGAGATC 314  
QY 101 CCTGCTACGACTGTGCGCCGACATGTGGCGGACTTTCTGCGAGGAAACCCCAACCTGATC 160  
315 CCTGTGGGAGTGCTCCAGGCGCATTTACGAAATTTTGTGGCGGATACCCCATGTACTC 374  
Db 161 TGAGATCTTCAACCGCGGCGCTCTA 185  
375 TGTATTATTATATAGCAGCGCTTTA 399  
QY 375 TGTATTATTATATAGCAGCGCTTTA 399  
Db 375 TGTATTATTATATAGCAGCGCTTTA 399

RESULT 7  
US-08-015-203-1  
Sequence 1, Application US/08015203  
Patent No. 5550034  
GENERAL INFORMATION:  
APPLICANT: Teng, Babie  
APPLICANT: Davidson, Nicholas O.  
APPLICANT: Burant, Charles F.  
TITLE OF INVENTION: Apolipoprotein B RNA Editing Protein:  
TITLE OF INVENTION: Composition and Method  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ARNOLD, WHITE & DURKEE  
STREET: 321 No. 5550034th Clark Street, Suite 800  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60610  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/015,203  
FILING DATE: 19930209  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Coolley, Ronald B.

REGISTRATION NUMBER: 27,187  
REFERENCE/DOCKET NUMBER: ARCD:069  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 744-0090  
TELEFAX: (312) 245-4861  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 879 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-015-203-1

Query Match  
Best Local Similarity 55.2%; Score 41; DB 1; Length 879;  
Matches 80; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

Db 41 CGAGCTGGAGCTTACCGCTGCGCGCTGCTACCGGCTGCTACCTGTTCACTCTGGAGCC 100  
255 CAGAAAGATACCTTTGTCTCAAAACACAGATGCTCATTAACCTGTTCTGTCGAGATC 314  
QY 101 CCTGCTACGACTGTGCGCCGACATGTGGCGGACTTTCTGCGAGGAAACCCCAACCTGATC 160  
315 CCTGTGGGAGTGCTCCAGGCGCATTTACGAAATTTTGTGGCGGATACCCCATGTACTC 374  
Db 161 TGAGATCTTCAACCGCGGCGCTCTA 185  
375 TGTATTATTATATAGCAGCGCTTTA 399  
QY 375 TGTATTATTATATAGCAGCGCTTTA 399  
Db 375 TGTATTATTATATAGCAGCGCTTTA 399

RESULT 8  
US-09-380-420C-1  
Sequence 1, Application US/09380420C  
Patent No. 6300544  
GENERAL INFORMATION:  
APPLICANT: Halkier, Barbara  
Halk, Soren  
Kahn, Rachel  
Moller, Birger  
TITLE OF INVENTION: Cytochrome P450 Monooxygenases  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Syngenta Patent Dept.  
STREET: 3054 Cornwalis Road  
CITY: RTP  
STATE: NC  
COUNTRY: USA  
ZIP: 27709  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/380,420C  
FILING DATE: 12-No. 6300544-1999  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Weiss, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: S-21251A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8587  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1929 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
IMMEDIATE SOURCE:  
CLONE: P450ox



FEATURE:  
NAME/KEY: CDS  
LOCATION: 81..1673  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-380-420C-1

Query Match 14.0%; Score 38; DB 4; Length 1929;  
Best Local Similarity 53.3%; Pred. No. 0.86;  
Matches 80; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 80 CCTGTTCACTTCCTGAGACCCCTGCTACGACTGTGCGGACATGTGGCCGACTTTCTGC 139  
DB 793 CCAGCTTCTCCGCGAGGACTTCTTCCCAACGCGCGCGGCGCTCGCCGACGCGCTCT 852  
QY 140 GAGGGAACCCCACTCAGTCTGAGATTTTCAACCGCGGCGCTTACTTCTGTAGAGAC 199  
DB 853 CGGGCTTCTCGCGCGCGCGCGGAGCGCATTTCAACGAGCTGACGCTTCTTGGAGAGG 912  
QY 200 GCAAGGCTGAGCCCGAGGGGCTGCGGCGGC 229  
DB 913 TCAATGACGACATGACATGACCGGCGCGCC 942

## RESULT 9

US-09-899-642A-1  
Sequence 1, Application US/09899642A  
Patent No. 6649814  
GENERAL INFORMATION:  
APPLICANT: Halkier, Barbara  
Bak, Soren

Kahn, Rachel  
Moller, Blitzer  
TITLE OF INVENTION: Cytochrome P450 Monooxygenases  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Syngenta Patent Dept.  
STREET: 3054 Cornwallis Road  
CITY: RTP  
STATE: NC  
COUNTRY: USA  
ZIP: 27709

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/899,642A  
FILING DATE: 05-Jul-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/380,420  
FILING DATE: 12-No. 6649814-1999  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: S-21251A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8587

## INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:  
LENGTH: 1929 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
IMMEDIATE SOURCE:  
CLONE: P450cx  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 81..1673  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-899-642A-1

Query Match 14.0%; Score 38; DB 4; Length 1929;  
Best Local Similarity 53.3%; Pred. No. 0.86;  
Matches 80; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 80 CCTGTTCACTTCCTGAGACCCCTGCTACGACTGTGCGGACATGTGGCCGACTTTCTGC 139  
DB 793 CCAGCTTCTCCGCGAGGACTTCTTCCCAACGCGCGCGGCGCTCGCCGACGCGCTCT 852  
QY 140 GAGGGAACCCCACTCAGTCTGAGATTTTCAACCGCGGCGCTTACTTCTGTAGAGAC 199  
DB 853 CGGGCTTCTCGCGCGCGCGGAGCGCATTTCAACGAGCTGACGCTTCTTGGAGAGG 912  
QY 200 GCAAGGCTGAGCCCGAGGGGCTGCGGCGGC 229  
DB 913 TCAATGACGACATGACATGACCGGCGCGCC 942

## RESULT 10

US-09-189-060B-73  
Sequence 73, Application US/09189060B  
Patent No. 6270968  
GENERAL INFORMATION:  
APPLICANT: Dalboge, Henrik  
Sandal, Thomas  
APPLICANT: Kauppinen, Markus  
Borge, Diderichsen

TITLE OF INVENTION: Method Of Providing No. 6270968el DNA Sequences  
FILE REFERENCE: 4772.204-US  
CURRENT APPLICATION NUMBER: US/09/189,060B  
PRIOR FILING DATE: 1998-11-10  
PRIOR APPLICATION NUMBER: PCT/DK97/00216  
PRIOR FILING DATE: 1997-05-12  
NUMBER OF SEQ ID NOS: 74  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 73  
LENGTH: 922  
TYPE: DNA  
ORGANISM: Hybrid

FEATURE:  
NAME/KEY: CDS

LOCATION: (2)...(922)  
US-09-189-060B-73

Query Match 13.9%; Score 37.8; DB 3; Length 922;  
Best Local Similarity 49.3%; Pred. No. 0.9;  
Matches 99; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 20 TGCCTTCTCCGCTACATCTCGGACTGGAGCTAGACCTTGCGCGCTGTAACCGGTCA 79  
DB 378 TGAATGCTCAGTGCACCAACGCGGCGGACTGTCCGACACCACTTGAGACTTCTCTCA 437  
QY 80 CCGGTTACCTCCTGAGACCCCTGCTACGACTGTGCGGACATGTGGCCGACTTTCTGC 139  
DB 438 TCCTCCGCGCGCGGCTGCGGATTTTCAACGCGCTGACCTTCCAGTACGCGCCCTGC 497  
QY 140 GAGGGAACCCCACTCAGTCTGAGATTTTCAACCGCGGCGCTTACTTCTGTAGAGAC 199  
DB 498 CGGGCGCGGAGTACGCGGCGGCTGAGCTCCGCGCGGAGTGCACGATGCGCGGAGCCA 557  
QY 200 GCAAGGCTGAGCCCGAGGGGCTGCGGCGGC 220  
DB 558 TCAAGCGCGGCTGCGGAGTGC 578

## RESULT 11

US-09-252-991A-10053/C  
Sequence 10053, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS



FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO: 10053  
LENGTH: 624  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-10053

Query Match 13.4%; Score 36.4; DB 4; Length 624;  
Best Local Similarity 51.9%; Pred. No. 1.9;  
Matches 82; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 87 CACCTCCCTGAGCCCTGCTACAGCTGTGCGGACATGTGGCCGACCTTTCTGCGAGGAA 146  
DB 367 CACTTCACTCTCTGAGCTGTGACCTTTACCCGCACTGAGGTGCGCTTGACCGGAT 308  
QY 147 CCCCACTCAGTGTGAGGATTTGACCGCGGCTTACTTCTGTGAGAACCGCAAGC 206  
DB 307 CGCTAACAGCGAGACACTATAGAGTGTGACAGATTCTTACAGCTCTGCACTC 248  
QY 207 TGAGCCCGAGGGGCTGCGGCGGCTGACCGCGCGG 244  
DB 247 GGAGCTGACAGACGTGCGCGCACCCGACGCGCAGAG 210

RESULT 12  
US-09-252-991A-9783  
Sequence 9783, Application US/09252991A  
Patent No. 6551735  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO: 9783  
LENGTH: 1959  
TYPE: DNA  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-9783

Query Match 13.4%; Score 36.4; DB 4; Length 1959;  
Best Local Similarity 51.9%; Pred. No. 2.1;  
Matches 82; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 87 CACCTCCCTGAGCCCTGCTACAGCTGTGCGGACATGTGGCCGACCTTTCTGCGAGGAA 146  
DB 1587 CACTTCACTCTCTGAGCTGTGACCTTTACCCGCACTGAGGTGCGCTTGACCGGAT 1646  
QY 147 CCCCACTCAGTGTGAGGATTTGACCGCGGCTTACTTCTGTGAGAACCGCAAGC 206  
DB 1647 CGCTAACAGCGAGACACTATAGAGTGTGACAGATTCTTACAGCTCTGCACTC 1706  
QY 207 TGAGCCCGAGGGGCTGCGGCGGCTGACCGCGCGG 244  
DB 1707 GGAGCTGACAGACGTGCGCGCACCCGACGCGCAGAG 1744

RESULT 13  
US-09-105-537-34  
Sequence 34, Application US/09105537A

Patent No. 6265202  
GENERAL INFORMATION:  
APPLICANT: Sherman, D.H.  
APPLICANT: Liu, H.  
APPLICANT: Xue, Y.  
APPLICANT: Xue, Y.  
TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
FILE REFERENCE: 600,438US1  
CURRENT APPLICATION NUMBER: US/09/105,537A  
CURRENT FILING DATE: 1998-06-26  
NUMBER OF SEQ ID NOS: 43  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO: 34  
LENGTH: 4689  
TYPE: DNA  
ORGANISM: Streptomyces venezuelae  
US-09-105-537-34

Query Match 13.2%; Score 35.8; DB 3; Length 4689;  
Best Local Similarity 46.9%; Pred. No. 3.3;  
Matches 112; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 4 GGCTGCACGTGGAATTGCTCTTCTTCTGCTACATCTGGAATGGAAGGACCTAGACCTTGGC 63  
DB 3406 GCCCACCACGAGCGGACCACTCTCTCTGCTACGCGGAGCGGAAAGCCCGCGA 3465  
QY 64 CGCTGTACCGGCTCACTGTCTTCACTTCTGAGACCTCTGTAGACTGTGCGCGACAT 123  
DB 3466 GCCACCACTACACCGCCGAACTACCGCATGGGCGCGCGGTACACATGCGCGCTGC 3525  
QY 124 GTGCGGACCTTCTGCGAGGAAACCCAACTCACTGATGAGATTTGACCGCGCGCTC 183  
DB 3526 GACGTGCGGACCTCCACCGCACTGCGACCTCTCTGAGGACATCCCGCGAGACGCG 3585  
QY 184 TACTTGTGAGAGCCGACGAGGCTGAGCCGAGGAGCTGCGGCGCTGACCGCGCGG 242  
DB 3586 CTACCGCGGTGTCCACACCGCGCGGCGGAGATCGCTGAGACTGACCGG 3644

RESULT 14  
US-09-105-537-5  
Sequence 5, Application US/09105537A  
Patent No. 6265202  
GENERAL INFORMATION:  
APPLICANT: Sherman, D.H.  
APPLICANT: Liu, H.  
APPLICANT: Xue, Y.  
APPLICANT: Xue, Y.  
TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
FILE REFERENCE: 600,438US1  
CURRENT APPLICATION NUMBER: US/09/105,537A  
CURRENT FILING DATE: 1998-06-26  
NUMBER OF SEQ ID NOS: 43  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO: 5  
LENGTH: 36778  
TYPE: DNA  
ORGANISM: Streptomyces venezuelae  
US-09-105-537-5

Query Match 13.2%; Score 35.8; DB 3; Length 36778;  
Best Local Similarity 46.9%; Pred. No. 4;  
Matches 112; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 4 GGCTGCACGTGGAATTGCTCTTCTGCTACATCTGGAATGGAAGGACCTAGACCTTGGC 63  
DB 30396 GCCCACCACGAGCGGACCACTCTCTCTGCTACGCGGAGCGGAAAGCCCGCGA 30455  
QY 64 CGCTGTACCGGCTCACTGTCTTCACTTCTGAGACCTCTGTAGACTGTGCGCGACAT 123  
DB 30456 GCCACCACTACCGCGGAACTGAGCGGCGCGGTACCATGCGCGCTGC 30515  
QY 124 GTGCGGACCTTCTGCGAGGAAACCCAACTCACTGAGATCTTACCGCGCGCTC 183



DB 30516 GACGTCGCCGACCCCGACCCATGCGCAACCTCTCGAGCCATCCCGCCGAGACGCC 30575  
QY 184 TACTTCTGTGAGGACCGGACGGCTGAGCCCGGCGGCTGCGCGGCTGCACCGCGCGG 242  
DB 30576 CTCACCGCGCTGTCACACCGCGCGGCGGCGGCGGATCGCTGACGTACCGG 30634

## RESULT 15

US-09-320-878-19

; Sequence 19, Application US/09320878A

; Patent No. 6117659

; GENERAL INFORMATION:

; APPLICANT: ASHLEY, Gary

; APPLICANT: BETLACH, Mary C.

; APPLICANT: MCDANIEL, Robert

; APPLICANT: TANG, Li

; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE

; FILE REFERENCE: 300622002120

; CURRENT APPLICATION NUMBER: US/09/320,878A

; CURRENT FILING DATE: 1999-05-27

; EARLIER APPLICATION NUMBER: CIP OF 09/141,908

; EARLIER FILING DATE: 1998-08-28

; EARLIER APPLICATION NUMBER: CIP OF 09/073,538

; EARLIER FILING DATE: 1998-05-06

; EARLIER APPLICATION NUMBER: CIP OF 08/846,247

; EARLIER FILING DATE: 1997-04-30

; EARLIER APPLICATION NUMBER: 60/119,139

; EARLIER FILING DATE: 1999-02-08

; EARLIER APPLICATION NUMBER: 60/100,880

; EARLIER FILING DATE: 1998-09-22

; EARLIER APPLICATION NUMBER: 60/087,080

; EARLIER FILING DATE: 1998-05-28

; NUMBER OF SEQ ID NOS: 34

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 19

; LENGTH: 38506

; TYPE: DNA

; ORGANISM: Streptomyces venezuelae

US-09-320-878-19

Query Match 13.2%; Score 35.8; DB 3; Length 38506;

Best Local Similarity 46.9%; Pred. No. 4;

Matches 112; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 4 GGCTGCGACGTGGAATTGCTTCTCCGCTACATCTGAGCTGGAGCTTACCTAGACCTGAC 63  
DB 28538 GCTCAGCAGGAGCCGACCACTCTCTCTCTGTCAGCCGAGGAGACAGCCCTCCGGA 28597  
QY 64 CGCTGTACCGCTCTCACTGTTCACTCTCTGAGCCCTGTACGACTGTGCGGACAT 123  
DB 28598 GCCACCAACTCACCGCCGACACTCACGCGATGCGGCGCCGCTACATGCGCGCTGC 28657  
QY 124 GTGCGCGACTTTTGTGAGGAGAACCCCAACTCACTGATCTGAGATCTTCAACCGCGCTC 183  
DB 28658 GACGTGCGGACCGCCGACCGCATGCGCACCTCTCTGAGCCATCCCGCCGAGAGGCC 28717  
QY 184 TACTTCTGTGAGGACCGGACGGCTGAGCCCGGCGGCTGCGCGGCTGCACCGCGCGG 242  
DB 28718 CTCACCGCGCTGTCACACCGCGGCGGCGGCGGCGGCGGATCGCTGAGCGTACCGG 28776

Search completed: March 13, 2004, 00:51:20  
Job time : 26.1637 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 12, 2004, 21:22:51 ; Search time 95.093 Seconds

(without alignments)  
10491.276 Million cell updates/sec

Title: US-09-966-880a-13

Sequence: 1 aacggctgcacagctggaattc.....agccatcatgaccttcaag 271

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 243257 seqs, 184079884 residues

Total number of hits satisfying chosen parameters: 4865114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
17: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*  
18: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	271	100.0	271	9	US-09-966-880a-13
2	271	100.0	2818	9	US-09-966-880a-7
3	271	100.0	6564	9	US-09-966-880a-10
4	271	100.0	11204	9	US-09-966-880a-35
5	221.6	81.8	2440	9	US-09-966-880a-1
6	87.4	32.3	2913	15	US-10-027-632-112023
7	87.4	32.3	2913	15	US-10-027-632-112024
8	87.4	32.3	20143	12	US-10-240-425-1099
9	85.8	31.7	987	9	US-09-925-300-699
10	85.8	31.7	1143	13	US-10-098-841-268
11	85	31.4	1155	15	US-10-460-923-1
12	85	31.4	1534	9	US-09-729-674-173
13	85	31.4	2151	9	US-09-822-830a-359
14	78.4	28.9	371	10	US-09-918-995-30237
15	76.8	28.3	823	14	US-10-247-671-62

16	76.8	28.3	1008	15	US-10-131-827-8890	Sequence 8890, Ap
17	76.8	28.3	1348	9	US-09-880-107-3282	Sequence 3282, Ap
18	68.4	25.2	456	14	US-10-029-386-25187	Sequence 25187, A
19	68.4	25.2	1404	14	US-10-198-846-13472	Sequence 13472, A
20	68.4	25.2	2773	15	US-10-104-047-1759	Sequence 1759, Ap
21	64.8	23.9	572	9	US-09-864-761-9416	Sequence 9416, Ap
22	63	23.2	476	10	US-09-918-985-32385	Sequence 32385, A
23	50.6	18.7	445	15	US-10-378-029-25	Sequence 25, Appl
24	50.6	18.7	879	14	US-10-157-031-13	Sequence 13, Appl
25	49.4	18.2	556	9	US-09-796-692-5061	Sequence 5061, Ap
26	49.4	18.2	556	14	US-10-040-862-5061	Sequence 5061, Ap
27	49.4	18.2	556	15	US-10-057-475B-5061	Sequence 5061, Ap
28	49.4	18.2	556	15	US-10-154-884B-5061	Sequence 5061, Ap
29	48	17.7	553	9	US-09-796-692-4504	Sequence 4504, Ap
30	48	17.7	553	14	US-10-040-862-4504	Sequence 4504, Ap
31	48	17.7	553	15	US-10-057-475B-4504	Sequence 4504, Ap
32	48	17.7	553	15	US-10-154-884B-4504	Sequence 4504, Ap
33	41	15.1	821	15	US-10-388-934-189	Sequence 189, Ap
34	40.6	15.0	807	12	US-10-424-599-135749	Sequence 135749, A
35	38.4	14.2	922	12	US-10-425-114-35403	Sequence 35403, A
36	38.4	14.2	1362	12	US-10-425-114-35704	Sequence 35704, A
37	38.2	14.1	60	10	US-09-908-978-7319	Sequence 7319, Ap
38	38	14.0	1929	10	US-09-899-642-1	Sequence 1, Appl1
39	37.2	13.7	1323	14	US-10-156-761-1176	Sequence 1176, Ap
40	37.2	13.7	9025608	14	US-10-156-761-1	Sequence 1, Appl1
41	37	13.7	520	14	US-10-184-634-332	Sequence 332, App
42	37	13.7	520	14	US-10-184-634-332	Sequence 332, App
43	37	13.7	674	14	US-10-257-826A-150	Sequence 150, App
44	37	13.7	1503	14	US-10-156-761-5170	Sequence 5170, Ap
45	37	13.7	9025608	14	US-10-156-761-1	Sequence 1, Appl1

## ALIGNMENTS

RESULT 1	
US-09-966-880a-13	
Sequence 13, Application US/09966880A	
Patent No. US2002016473A1	
GENERAL INFORMATION:	
APPLICANT: Honjo, Tasuku	
APPLICANT: Muramatsu, Masamichi	
TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE	
FILE REFERENCE: 06501-088001	
CURRENT APPLICATION NUMBER: US/09/966, 880A	
CURRENT FILING DATE: 2001-09-28	
PRIOR APPLICATION NUMBER: PCT/JP00/01918	
PRIOR FILING DATE: 2000-03-28	
PRIOR APPLICATION NUMBER: JP 11-371382	
PRIOR FILING DATE: 1999-12-27	
PRIOR APPLICATION NUMBER: JP 11-178999	
PRIOR FILING DATE: 1999-06-24	
PRIOR APPLICATION NUMBER: JP 11-87192	
PRIOR FILING DATE: 1999-03-29	
NUMBER OF SEQ ID NOS: 36	
SOFTWARE: FastSeq for Windows Version 4.0	
SEQ ID NO 13	
LENGTH: 271	
TYPE: DNA	
ORGANISM: Homo sapiens	
US-09-966-880a-13	
Query Match	
Best Local Similarity 100.0%; Score 271; DB 9; Length 271;	
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Cy 1 AACGGCTGCCACGATGATTCTCTTCTCCGCTACATCTTGAGCTGGAGCTAGACCT 60	
Db 1 AACGGCTGCCACGATGATTCTCTTCTCCGCTACATCTTGAGCTGGAGCTAGACCT 60	
Cy 61 GAGCGCTGCTACCGCCGCTACCTGTTCACTCTCTGAGCCCGCTACGACTGTGCCGA 120	
Db 61 GAGCGCTGCTACCGCCGCTACCTGTTCACTCTCTGAGCCCGCTACGACTGTGCCGA 120	



```

QY 121 CATGTCGCGGACTTCTGTGAGGAGAACCCCAACTCAGTCTGAGGATCTTCAACCGCGGC 180
Db 121 CATGTCGCGGACTTCTGTGAGGAGAACCCCAACTCAGTCTGAGGATCTTCAACCGCGGC 180
QY 181 CTCTACTTCTGTGAGGAGCGCAAGGCTGAGCCCGAGGAGGCTGCGCGGCTGACCGCGCC 240
Db 181 CTCTACTTCTGTGAGGAGCGCAAGGCTGAGCCCGAGGAGGCTGCGCGGCTGACCGCGCC 240
QY 241 GGGGTGCAAAATAGCCATCATGACCTTCAAG 271
Db 241 GGGGTGCAAAATAGCCATCATGACCTTCAAG 271

RESULT 2
US-09-966-880A-7
; Sequence 7, Application US/09966880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; APPLICANT: Muramatsu, Masamichi
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966,880A
; PRIOR FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 2818
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (80)...(673)
; FEATURE:
; NAME/KEY: 5' UTR
; LOCATION: (1)...(79)
; FEATURE:
; NAME/KEY: 3' UTR
; LOCATION: (677)...(2818)
; US-09-966-880A-7

Query Match 100.0%; Score 271; DB 9; Length 2818;
Best Local Similarity 100.0%; Pred. No. 1.9e-68;
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

RESULT 3
US-09-966-880A-10
; Sequence 10, Application US/09966880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; APPLICANT: Muramatsu, Masamichi
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966,880A
; PRIOR FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 6564
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-966-880A-10

Query Match 100.0%; Score 271; DB 9; Length 6564;
Best Local Similarity 100.0%; Pred. No. 1.9e-68;
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```



NUMBER OF SEQ ID NOS: 36  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 35  
 LENGTH: 11204  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 US-09-966-880A-35

Query Match 100.0%; Score 271; DB 9; Length 11204;  
 Best Local Similarity 100.0%; Pred. No. 2e-68;  
 Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACGGCTCCAGCGGAAATGCTCTCTCCGCTACATCTGAGACTGGAGCTAGACCT 60  
 DB 7807 AACGGCTCCAGCGGAAATGCTCTCTCCGCTACATCTGAGACTGGAGCTAGACCT 7866  
 QY 61 GAGCGCTGCTACCGCGTCACTGCTTCACTCTGAGCCCTGCTACGACTGTGCCGA 120  
 DB 7867 GAGCGCTGCTACCGCGTCACTGCTTCACTCTGAGCCCTGCTACGACTGTGCCGA 7926  
 QY 121 CATGTGCGGACTTCTTCTGAGAGGAAACCCCACTCACTGAGATCTTCAACCGCGC 180  
 DB 7927 CATGTGCGGACTTCTTCTGAGAGGAAACCCCACTCACTGAGATCTTCAACCGCGC 7986  
 QY 181 CTCTACTTCTGTGAGAGACCGCAAGCTGAGCCGAGCGGCTGCGGCTGCAACCGGCG 240  
 DB 7987 CTCTACTTCTGTGAGAGACCGCAAGCTGAGCCGAGCGGCTGCGGCTGCAACCGGCG 8046  
 QY 241 GGGGTGCAATPAGCATCATGACCTTCAAG 271  
 DB 8047 GGGGTGCAATPAGCATCATGACCTTCAAG 8077

## RESULT 5

US-09-966-880A-1  
 Sequence 1, Application US/09966880A  
 Patent No. US20020164743A1  
 GENERAL INFORMATION:  
 APPLICANT: Honjo, Tasuku  
 APPLICANT: Muramatsu, Masaamichi  
 TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
 FILE REFERENCE: 06501-088001  
 CURRENT APPLICATION NUMBER: US/09/966.880A  
 PRIOR FILING DATE: 2001-09-28  
 PRIOR APPLICATION NUMBER: PCT/JP00/01918  
 PRIOR FILING DATE: 2000-03-28  
 PRIOR APPLICATION NUMBER: JP 11-371382  
 PRIOR FILING DATE: 1999-12-27  
 PRIOR APPLICATION NUMBER: JP 11-178999  
 PRIOR FILING DATE: 1999-06-24  
 PRIOR APPLICATION NUMBER: JP 11-87192  
 PRIOR FILING DATE: 1999-03-29  
 NUMBER OF SEQ ID NOS: 36  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 1  
 LENGTH: 2440  
 TYPE: DNA  
 ORGANISM: Mus musculus  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (93)...(686)  
 FEATURE:  
 NAME/KEY: 5'UTR  
 LOCATION: (1)...(92)  
 FEATURE:  
 NAME/KEY: 3'UTR  
 LOCATION: (690)...(2440)  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: (1)...(2440)  
 OTHER INFORMATION: n = A,T,C or G  
 US-09-966-880A-1

Query Match 81.8%; Score 221.6; DB 9; Length 2440;  
 Best Local Similarity 89.2%; Pred. No. 3e-54;  
 Matches 239; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 4 GGGTGGCAAGTGAATTTGCTCTTCTCCGCTACATCTGAGACTGGAGCTTAGACCTTGGC 63  
 DB 252 GGGTGGCAAGTGAATTTGCTCTTCTCCGCTACATCTGAGACTGGAGCTTAGACCTTGGC 311  
 QY 64 CGGTGTACCGGCTACCGGCTGCTCACTCTGAGCCCTGCTACGACTGTGCCGACAT 123  
 DB 312 CGGTGTACCGGCTACCGGCTGCTCACTCTGAGCCCTGCTACGACTGTGCCGACAT 371  
 QY 124 GTGAGCGGACTTCTGAGAGGAAACCCCACTCACTGAGATCTTCAACCGCGCGCTC 183  
 DB 372 GTGAGCGGACTTCTGAGAGGAAACCCCACTCACTGAGATCTTCAACCGCGCGCTC 431  
 QY 184 TACTTCTGTGAGAGACCGCAAGGCTGAGCCGAGGCTGCGGCGGCTGACCGCGCGG 243  
 DB 432 TACTTCTGTGAGAGACCGCAAGGCTGAGCCGAGGCTGCGGCGGCTGACCGCGCGG 491  
 QY 244 GTGCAATPAGCATCATGACCTTCAAG 271  
 DB 492 GTGCAATPAGCATCATGACCTTCAAG 519

## RESULT 6

US-10-027-632-112023  
 Sequence 112023, Application US/10027632  
 Publication No. US20030204075A9  
 GENERAL INFORMATION:  
 APPLICANT: Mang, David G.  
 TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
 POLYMORPHISMS IN THE HUMAN GENOME  
 FILE REFERENCE: 108827.129  
 CURRENT APPLICATION NUMBER: US/10/027.632  
 CURRENT FILING DATE: 2002-04-30  
 PRIOR APPLICATION NUMBER: US 60/218,006  
 PRIOR FILING DATE: 2000-07-12  
 PRIOR APPLICATION NUMBER: US 60/198,676  
 PRIOR FILING DATE: 2000-04-20  
 PRIOR APPLICATION NUMBER: US 60/193,483  
 PRIOR FILING DATE: 2000-03-29  
 PRIOR APPLICATION NUMBER: US 60/185,218  
 PRIOR FILING DATE: 2000-02-24  
 PRIOR APPLICATION NUMBER: US 60/167,363  
 PRIOR FILING DATE: 1999-11-23  
 PRIOR APPLICATION NUMBER: US 60/156,358  
 PRIOR FILING DATE: 1999-09-28  
 PRIOR APPLICATION NUMBER: US 60/146,002  
 PRIOR FILING DATE: 1999-08-09  
 NUMBER OF SEQ ID NOS: 325720  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 112023  
 LENGTH: 2913  
 TYPE: DNA  
 ORGANISM: Human  
 US-10-027-632-112023

Query Match 32.3%; Score 87.4; DB 15; Length 2913;  
 Best Local Similarity 60.8%; Pred. No. 1.2e-15;  
 Matches 161; Conservative 0; Mismatches 101; Indels 3; Gaps 1;

QY 7 TGGCAGGTGAATTTGCTCTTCTCCGCTACATCTGAGACTGGAGCTTAGACCTTGGCGCC 66  
 DB 11 TGGCAGGTGAATTTGCTCTTCTCCGCTACATCTGAGACTGGAGCTTAGACCTTGGCGCC 70  
 QY 67 TGGTACCGGCTACCGGCTGCTCACTCTGAGAGCCCTGCTACGACTGTGCCGACATGTG 126  
 DB 71 TGGTACCGGCTACCGGCTGCTCACTCTGAGAGCCCTGCTACGACTGTGCCGACATGTG 130  
 QY 127 GCGGACTTCTGAGAGGAAACCCCACTCACTGAGATCTTCAACCGCGGCGCTTAC 186  
 DB 131 GCGGACTTCTGAGAGGAAACCCCACTCACTGAGATCTTCAACCGCGGCGCTTAC 190



QY 187 TTCTGTGAGGACCGGAGGCTGAGCCGAGGGGCTGGGGGGCTGACACCGGGCGGGGCTG 246  
DB 191 TACTTCTGGGA---TACAGATTACAGAGGGGCTCCGACCTGAGTCAAGAAAGGGGCC 247  
QY 247 CAATAGCCATCATGACCTTCAAG 271  
DB 248 TCCGTGAGATCATGGGCTACAAAG 272

## RESULT 7

US-10-027-632-112024  
Sequence 112024, Application US/10027632  
Publication No. US20030204075A9  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
POLYMORPHISMS IN THE HUMAN GENOME  
FILE REFERENCE: 108827.129  
CURRENT FILING DATE: 2002-04-30  
PRIOR APPLICATION NUMBER: US 60/218,006  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/198,676  
PRIOR FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: US 60/193,483  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: US 60/185,218  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/167,363  
PRIOR FILING DATE: 1999-11-23  
PRIOR APPLICATION NUMBER: US 60/156,358  
PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: US 60/146,002  
PRIOR FILING DATE: 1999-08-09  
NUMBER OF SEQ ID NOS: 325720  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 112024  
LENGTH: 2913  
TYPE: DNA  
ORGANISM: Human  
US-10-027-632-112024

Query Match 32.3%; Score 87.4; DB 15; Length 2913;  
Best Local Similarity 60.8%; Pred. No. 1.2e-15;  
Matches 161; Conservative 0; Mismatches 101; Indels 3; Gaps 1;

QY 7 TCCACGCTGGAATTGCTCTTCTCCGCTACATCTGGACTGGAGCTGAGCTTGGAGCCG 66  
DB 11 TGTCAATGACGAAGGCTCTTCTCTTGTGTGACGACATCTCTCTTAACACA 70  
QY 67 TCGTACCGGCTACCTGCTTCACTCTGAGAGCCCTGCTAGACTGTGCGGACATGTG 126  
DB 71 AACTAGAGGTACCTGCTGACATCTTGAAGCCCTTGCAGAGTGTGAGGGAGGTG 130  
QY 127 GCCGACTTCTGCGAGGAGAACCCCACTCACTGAGAGATTTTCAACCGCGGCTTAC 186  
DB 131 GCCGAGTCTGCGGAGGACGACGAGATTCACATTTTCAACCGCGGCTTAC 190  
QY 187 TTCTGTGAGGACCGCAAGCTGAGCCGAGGGGCTGCGCGGCTGACCGCGCGGGTG 246  
DB 191 TACTTCTGGGA---TACAGATTACAGAGGGGCTCCGAGCTGAGTCAAGAAAGGGGCC 247  
QY 247 CAATAGCCATCATGACCTTCAAG 271  
DB 248 TCCGTGAGATCATGGGCTACAAAG 272

RESULT 8  
US-10-240-425-1099  
Sequence 1099, Application US/10240425  
Publication No. US20040033502A1  
GENERAL INFORMATION:

APPLICANT: Williams, Amanda  
APPLICANT: Boland, Joseph F.  
APPLICANT: Lord, Reginald V.  
APPLICANT: Alvarez, Chris  
APPLICANT: Wetzel, Jon C.  
APPLICANT: Schert, Uwe  
APPLICANT: Vockley, Joseph G.  
TITLE OF INVENTION: Gene Expression Profiles in Esophageal Tissue  
FILE REFERENCE: 44921-5026  
CURRENT FILING DATE: 2002-09-30  
PRIOR APPLICATION NUMBER: PCT/US01/09847  
PRIOR FILING DATE: 2001-03-28  
PRIOR APPLICATION NUMBER: US 60/193,446  
PRIOR FILING DATE: 2000-03-31  
NUMBER OF SEQ ID NOS: 1588  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1099  
LENGTH: 201143  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: Genbank Accession No. US20040033502A1 AL022318  
US-10-240-425-1099

Query Match 32.3%; Score 87.4; DB 12; Length 201143;  
Best Local Similarity 60.8%; Pred. No. 1.4e-15;  
Matches 161; Conservative 0; Mismatches 101; Indels 3; Gaps 1;

QY 7 TCCACGCTGGAATTGCTCTTCTCCGCTACATCTGGAAGCTGGAGCTGAGCTTGGAGCCG 66  
DB 167899 TGTCAATGACGAAGGCTCTTCTCTTGTGTGACGACATCTCTCTTAACACA 167958  
QY 67 TCGTACCGGCTACCTGCTTCACTCTGAGAGCCCTGCTAGACTGTGCGGACATGTG 126  
DB 167959 AACTAGAGGTACCTGCTGACATCTTGAAGCCCTTGCAGAGTGTGAGGGAGGTG 168018  
QY 127 GCCGACTTCTGCGAGGAGAACCCCACTCACTGAGAGATTTTCAACCGCGGCTTAC 186  
DB 168019 GCCGAGTCTGCGGAGGACGACGAGATTCACATTTTCAACCGCGGCTTAC 168078  
QY 187 TTCTGTGAGGACCGCAAGCTGAGCCGAGGGGCTGCGCGGCTGACCGCGCGGGTG 246  
DB 168079 TACTTCTGGGA---TACAGATTACAGAGGGGCTCCGAGCTGAGTCAAGAAAGGGGCC 168135  
QY 247 CAATAGCCATCATGACCTTCAAG 271  
DB 168136 TCCGTGAGATCATGGGCTACAAAG 168160

## RESULT 9

US-09-925-300-699  
Sequence 699, Application US/09925300  
Patent No. US20020151681A1  
GENERAL INFORMATION:  
APPLICANT: Craig Rosen,  
APPLICANT: Steve Ruben,  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA101  
CURRENT FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05988  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 1690  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 699  
LENGTH: 987  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-925-300-699



Query Match	31.7%;	Score 85.8;	DB 9;	Length 987;
Best Local Similarity	60.4%;	Pred. No. 3.4e-15;		
Matches 160;	Conservative 0;	Mismatches 102;	Indels 3;	Gaps 1;

Qy	TGCAAGTGAATTCCTCCTCCCGGCTACATCTGGACTGGAACTGAAACCCTGGCGGC	66
Db	291 TGTCAATGCAAGAAGTGCTTCTCTCTCTTGTTGTTCTGGACGACATATCTCTCTTAACA	350
Qy	67 TGCTACCGCGTCACTGTTCACCTTCCTGGACCCCTGCTACGACTGTGCCGACATGTG	126
Db	351 AAGTACAGAGTCACTGGTACCATTTTGAACTCTTCCAGACTGTGCAGGGAGATG	410
Qy	127 GCGAATTTCGAGAGGAACCCCAAACCTCACTGAGATCTTCAACCGCGCGCTTAC	186
Db	411 GCCGAGTTCCTGGCCAGGACAGAACGTGAATCTCACCACTTTCACCGCCCGCTTAC	470
Qy	187 TTCTGTGAGGACCGCAAGGCTGAGCCCGAGGAGCTTGCGAGGCTGCACCCGCGCCGGGTG	246
Db	471 TAATTCAGATAT---CATGTTACCGAGAGGGGGCTCCGACGCTGATGACGAGAGGGGTC	527
Qy	247 CAATGGCATCATGACCTTCAAG	271
Db	528 GCTGTGAGATCATGACATATGAAAG	552

RESULT 10  
US-10-098

```

? Sequence 268, Application US/10038841
? Publication No. US20020197679A1
? GENERAL INFORMATION:
? APPLICANT: Tang, Y. Tom
? APPLICANT: Liu, Chenghua
? APPLICANT: Asundi, Vinod
? APPLICANT: Xu, Chongjun
? APPLICANT: Zhou, Ping
? APPLICANT: Ma, Yundong
? APPLICANT: Wang, Jian-Rui
? APPLICANT: Zhao, Qing A.
? APPLICANT: Ren, Feiyao
? APPLICANT: Chen, Rui-hong
? APPLICANT: Wang, Dunrui
? APPLICANT: Wang, Zhwei
? APPLICANT: Weinman, Tom
? APPLICANT: Zhang, Jie
? APPLICANT: Qian, Xiaohong B.
? APPLICANT: Dimana, Radjoie T.
? TITLE OF INVENTION: No. US20020197679A1 Nuclear Acids and
? TITLE OF INVENTION: Polypeptides
? FILE REFERENCE: 784C1P2
? CURRENT APPLICATION NUMBER: US/10/098,841
? CURRENT FILING DATE: 2002-03-13
? PRIOR APPLICATION NUMBER: 09/598,042
? PRIOR FILING DATE: 2000-06-20
? PRIOR APPLICATION NUMBER: 09/552,317
? PRIOR FILING DATE: 2000-04-25
? PRIOR APPLICATION NUMBER: 09/488,725
? PRIOR FILING DATE: 2000-01-21
? NUMBER OF SEQ ID NOS: 331
? SOFTWARE: pl_FL_genes Version 1.0
? SEQ ID NO 268
? LENGTH: 1143
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURES:
? NAME/KEY: CDS
? LOCATION: (128)..(700)
? US-10-098-841-268

```

Query	7	TGCCACGTGGAAATTGCTCTTTCTCCGCATATCGGAGATGGGAACTAGACCTGGCGCG	66
Query Match	31.7%	Score 85.8; DB 13;	Length 1143;
Best Local Similarity	60.4%;	Pred. No. 3,4e-15;	
Matches 160; Conservative	0;	Mismatches 102;	Indels 3; Gaps 1

7 TGCACGTTGCTCTTCCCTACATCTGGACTGGACCTAGACCTGGCCG 66

Db	320	TGTCATGACGAAAGGCTTCCTCCTCTTGATTCTGCGACGACATACTGTCTCTCAACACA	379
Qy	67	TGTCACCGGGTACCCGAGTTACCTCCCTCGAGCCCTCTACACTGTGCGCCGACATGTG	126
Db	380	AAGTACCAAGTCACTCGTGTACATCTTGGACCCTTGCCAGACTGTGGACGGGAGGTG	439
Qy	127	GCCGACTTCTGCGAGGGAAACCCCACTCACTGTGAGGATCTTTCACCCGGCGCTTAC	186
Db	440	GCCGAGTTCCTGCCGACGACGACAGCAAGTGAATCTCAACATCTTTCACCGCCGCTTAC	499
Qy	187	TTCTGTGAGGACCGCGAAGGCTGAGGCCGAGGGGCTGCGCGGCTGCAACCGCGCCGGGTG	246
Db	500	TACTTCAAGATC---CATGTTACAGAGGGGGCTCCGCACTGAGTCAAGAGGAGGTG	556
Qy	247	CAATAGCCATCATGACCTTCAAG	271
Db	557	GCTGTGAGATCATGACTATGAG	581

RESULT 11  
US-10-460

```

; Sequence 1, Application US/10460923
; Publication No. US2004000951a1
;
GENERAL INFORMATION:
;
APPLICANT: MALIM, Michael H.
APPLICANT: SHERBY, Ann M.
APPLICANT: HARRIS, Reuben S.
APPLICANT: BISHOP, Kate N.
APPLICANT: NEUBERGER, Michael S.
APPLICANT: GADDIS, Nathan C.
APPLICANT: SIMON, James H.M.
TITLE OF INVENTION: DNA Deamination Mediates Innate Immunity to Retroviral Infection
FILE REFERENCE: 22253-74380
CURRENT APPLICATION NUMBER: US/10/460,923
CURRENT FILING DATE: 2003-06-13
PRIOR APPLICATION NUMBER: US 60/388,513
PRIOR FILING DATE: 2002-06-13
PRIOR APPLICATION NUMBER: US 60/472,952
PRIOR FILING DATE: 2003-05-23
NUMBER OF SEQ ID NOS: 12
SOFTWARE: Patentin version 3.1
;
SEQ ID NO 1
;
LENGTH: 1155
;
TYPE: DNA
;
ORGANISM: Homo sapiens
;
US-10-460-923-1

```

Query March	31.4%;	Score 85;	DB 15;	Length 1155;
Best Local Similarity	60.3%;	Pred. No. 5.7e-15;		
Matches 161; Conservative	0;	Mismatches 100;	Indels 6;	Gaps 1;

QY	2	ACGCGTCGACGCGAAATGCTCTTCCCTCCGCTACATCTCGAATGGGACCTTAGCCCTGG	61
DB	761	AAAGCCGCCATGAGAGCTGTGCTTCCTCGAAGGTGATTCCTTTTGGAACTGGACTGG	820
QY	62	GCCGCTGTAACGCGCTCACTGGTTCACCTCTCTGGAGCCCTGTTACGACTGTGCTCCGAC	121
DB	821	ACGAGGACTACAGGGGTATTACTGTCTTCACTCTGTGAGGCCCTGTTCAGCTGTGTGCCAGG	880
QY	122	ATGCGGCCGACTTTCGTGGAGGGGAAACCCCAACTCAAGTCTGAGGATCTTCAACGCGGCGC	181
DB	881	AAATGGCTAAATTCATTTTCAAAAAACAAACAGTGAAGCTGTGATTTTCACTGCCGCA	940
QY	182	TCTACTTCTGTAGGACCGCAAGGCTGAGGCCGCAAGGGCTGCGGCGGCTGCACCGCGCG	241
DB	941	TCTA-----TGATGATCAAGGAAGATGTCAAGAGGGGCTGCGCACCTGGCGAGGCTG	994
QY	242	GGGTGCAATAGCATCATGACCTTCA	266
DB	995	GGGCGCAAAATTCATAATAGACATACA	1021



## RESULT 12

US-09-729-674-173

Sequence 173, Application US/09729674

Patent No. US2001003935A1

GENERAL INFORMATION:

APPLICANT: Jacobs, Kenneth  
APPLICANT: McCoy, John M.  
APPLICANT: LaValle, Edward R.  
APPLICANT: Collins-Racie, Lisa A.  
APPLICANT: Evans, Cheryl  
APPLICANT: Merberg, David  
APPLICANT: Treacy, Maurice  
APPLICANT: Agostino, Michael J.  
APPLICANT: Steindinger II, Robert J.  
APPLICANT: Spaulding, Vikki  
APPLICANT: Wong, Gordon G.  
APPLICANT: Clark, Hilary  
APPLICANT: Rechele, Kim  
APPLICANT: Genetics Institute, Inc.  
FILE REFERENCE: 6055-64X  
CURRENT APPLICATION NUMBER: US/09/729,674  
CURRENT FILING DATE: 2000-12-04  
PRIOR APPLICATION NUMBER: 09/539,330  
PRIOR FILING DATE: 2000-03-30  
NUMBER OF SEQ ID NOS: 283  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 173  
LENGTH: 1534  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-729-674-173

Query Match 31.4%; Score 85; DB 9; Length 1534;  
Best Local Similarity 60.3%; Pred. No. 5.8e-15;  
Matches 161; Conservative 0; Mismatches 100; Indels 6; Gaps 1;

QY 2 ACGGCTGCCAGTGGAAATGCTCTCCCTGCTACATCTCGAGAGGAGACCTAGACCTG 61  
DB 811 AAGGCCCGCATGACAGCTGTCTCTCTGAGACGTATTCCTTTGGAAGCTGGACCTGG 870  
QY 62 GCCGCTGTACCGCGTCACTGCTTCACTCCTCTGAGCCCTGCTACGACTGTGCCGAC 121  
DB 871 ACCAGACTACAGGGTTACTCTCTCACTCCTCTGAGCCCTGCTTCACTGTGCCAG 930  
QY 122 ATGTGGCCGACTTTCTGCGAGGAAACCCCACTGATGTGAGATCTTCAACCGGCGCC 181  
DB 931 AATGGCTAAATTCATTTCAAAAACAAACAGTGAAGCTGTGATCTTCACTGCCGCA 990  
QY 182 TCTACTCTGTGAGACCGCAAGCGCTGAGCCCGAGGGGCTGGGGGCTGACACCGCGCG 241  
DB 991 TCTA-----TGATATCAAGGAAGATGTCAAGAGGGGCTGGCAACCTGGCCGAGCTG 1044  
QY 242 GGGTGCATAATAGCATCATGACCTTCA 268  
DB 1045 GGGCCAAATTTCAATATGACATACA 1071

## RESULT 13

US-09-822-830A-359

Sequence 359, Application US/09822830A

Patent No. US20020142952A1

GENERAL INFORMATION:

APPLICANT: Genetics Institute, Inc.  
APPLICANT: Wong, Gordon G.  
APPLICANT: Clark, Hilary  
APPLICANT: Rechele, Kim  
APPLICANT: Agostino, Michael J.  
APPLICANT: Howes, Steven H.  
APPLICANT: Resnick, Richard J.  
APPLICANT: Gulukota, Kamalakhar  
APPLICANT: Graham, James R.  
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS

FILE REFERENCE: GIN 6402  
CURRENT APPLICATION NUMBER: US/09/822,830A  
CURRENT FILING DATE: 2001-03-29  
PRIOR APPLICATION NUMBER: 60/195,604  
PRIOR FILING DATE: 2000-04-06  
NUMBER OF SEQ ID NOS: 631  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 359  
LENGTH: 2151  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-822-830A-359

Query Match 31.4%; Score 85; DB 9; Length 2151;  
Best Local Similarity 60.3%; Pred. No. 5.9e-15;  
Matches 161; Conservative 0; Mismatches 100; Indels 6; Gaps 1;

QY 2 ACGGCTGCCAGTGGAAATGCTCTCCCTGCTACATCTCGAGAGGAGACCTAGACCTG 61  
DB 1450 AAGGCCCGCATGACAGCTGTCTCTCTGAGACGTATTCCTTTGGAAGCTGGACCTGG 1509  
QY 62 GCCGCTGTACCGCGTCACTGCTTCACTCCTCTGAGCCCTGCTACGACTGTGCCGAC 121  
DB 1510 ACCAGACTACAGGGTTACTCTCTCACTCCTCTGAGCCCTGCTTCACTGTGCCAG 1569  
QY 122 ATGTGGCCGACTTTCTGCGAGGAAACCCCACTGATGTGAGATCTTCAACCGGCGCC 181  
DB 1570 AATGGCTAAATTCATTTCAAAAACAAACAGTGAAGCTGTGATCTTCACTGCCGCA 1629  
QY 182 TCTACTCTGTGAGACCGCAAGCGCTGAGCCCGAGGGGCTGGGGGCTGACACCGCGCG 241  
DB 1630 TCTA-----TGATATCAAGGAAGATGTCAAGAGGGGCTGGCAACCTGGCCGAGCTG 1683  
QY 242 GGGTGCATAATAGCATCATGACCTTCA 268  
DB 1684 GGGCCAAATTTCAATATGACATACA 1710

## RESULT 14

US-09-918-995-30237

Sequence 30237, Application US/09918995

Publication No. US20030073623A1

GENERAL INFORMATION:

APPLICANT: Hyseq, Inc.  
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
FILE REFERENCE: 20411-756  
CURRENT APPLICATION NUMBER: US/09/918,995  
CURRENT FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: US/09/235,076  
PRIOR FILING DATE: 1999-01-20  
NUMBER OF SEQ ID NOS: 38054  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 30237  
LENGTH: 371  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-918-995-30237

Query Match 28.9%; Score 78.4; DB 10; Length 371;  
Best Local Similarity 64.1%; Pred. No. 4.4e-13;  
Matches 118; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 2 ACGGCTGCCAGTGGAAATGCTCTCCCTGCTACATCTCGAGAGGAGACCTAGACCTG 61  
DB 171 AAGGCCCGCATGACAGCTGTCTCTCTGAGACGTATTCCTTTGGAAGCTGGACCTGG 230  
QY 62 GCCGCTGTACCGCGTCACTGCTTCACTCCTCTGAGCCCTGCTACGACTGTGCCGAC 121  
DB 231 ACCAGACTACAGGGTTACTCTCTCACTCCTCTGAGCCCTGCTTCACTGTGCCAG 290  
QY 122 ATGTGGCCGACTTTCTGCGAGGAAACCCCACTGATGTGAGATCTTCAACCGGCGCC 181



Db 291 AAATGCTAAATTCATTTCAAAAAACAACAGTGAGCCTGTGATCTTCACTGCGCGCA 350  
QY 182 TCTA 185  
182 TCTA 185  
Db 351 TCTA 354

## RESULT 15

US-10-247-671-62  
; Sequence 62, Application US/10247671  
; Publication No. US20030194721A1  
; GENERAL INFORMATION:  
; APPLICANT: Mikita, Thomas  
; APPLICANT: Shiffman, Dov  
; APPLICANT: Porter, Gordon, J.  
; APPLICANT: Kaser, Matthew R.  
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED FOAM CELLS  
; FILE REFERENCE: PA-0050 US  
; CURRENT APPLICATION NUMBER: US/10/247,671  
; CURRENT FILING DATE: 2002-09-18  
; PRIOR APPLICATION NUMBER: 60/323,784  
; PRIOR FILING DATE: 2001-09-19  
; NUMBER OF SEQ ID NOS: 186  
; SOFTWARE: PERL Program  
; SEQ ID NO 62  
; LENGTH: 823  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. US20030194721A1 135626CB1  
US-10-247-671-62

Query Match 28.3%; Score 76.8; DB 14; Length 823;  
Best Local Similarity 60.5%; Pred. No. 1.3e-12;  
Matches 167; Conservative 0; Mismatches 97; Indels 12; Gaps 2;

QY 2 ACGGCTGCCACGTGGAATTCCTTCCTCCGCTACATCTGGAAGTGGAGCTAGACCTG 61  
Db 115 ACGGCGCCATGCGAGCTGCGCTTCTTGACCTGCTTCTTGTGAGTTGAGACCCGG 174  
QY 62 GCGGCTGCTAACGCGCTGACCTGCTGAGAGCCCGCTAC-----GACTG 115  
Db 175 CCGAGTCTACAGGGTCACTTGCTGATCTCTGAGCCCTGCTTCTTCTGAGGCTG 234  
QY 116 CCGACATGTGGCGGCACTTCTGCGAGGAGACCCCACTCACTGAGATCTTCAACG 175  
Db 235 CCGGGGAAGTGCCTGCTTCTTCTGAGAGACACACACCTGAGACTGCGCATCTTGG 294  
QY 176 CCGGCTCTACTTCTGTGAGAGACCGAGAGCTGAGCCCGAGGCGCTGCGCGCTG 235  
Db 295 CCGGCACTATGATGATGACCC-----TATATAGAGAGCGCTGCAATGCTGCGG 348  
QY 236 GCGCCGCGGCTGCAATAGCATGATGACCTTCAAG 271  
Db 349 ATGCTGGGGCCCAAGTCTCATCATGACCTACGATG 384

Search completed: March 13, 2004, 05:31:06  
Job time: 105.099 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 12, 2004, 13:19:34 ; Search time 862.933 Seconds

(without alignments)  
9378.080 Million cell updates/sec

Title: US-09-966-880a-13

Perfect score: 271  
Sequence: 1 aacggctgccacgtggaatt.....agccatcatgaccttcaag 271

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_estba:\*  
3: em\_estba:\*  
4: em\_estba:\*  
5: em\_estba:\*  
6: em\_estba:\*  
7: em\_estba:\*  
8: em\_estba:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_est3:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estba:\*  
16: em\_estba:\*  
17: em\_estba:\*  
18: em\_estba:\*  
19: em\_estba:\*  
20: em\_estba:\*  
21: em\_estba:\*  
22: em\_estba:\*  
23: em\_estba:\*  
24: em\_estba:\*  
25: em\_estba:\*  
26: em\_estba:\*  
27: em\_estba:\*  
28: gb\_est1:\*  
29: gb\_est2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	271	100.0	541	10	BF238155 601811880
2	271	100.0	693	12	BG757392 602711022
3	271	100.0	743	12	BG686133 602638412
4	271	100.0	820	12	BG757089 602715124

5	271	100.0	853	13	BK64579	BK64579
6	271	100.0	872	12	BG758510	BG758510
7	271	100.0	889	12	BG686876	BG686876
8	271	100.0	953	13	BQ065440	BQ065440
9	271	100.0	1052	13	BQ055935	BQ055935
10	271	100.0	1201	9	AL559877	AL559877
11	271	100.0	1201	13	BK402063	BK402063
12	269.4	99.4	942	10	BF975166	BF975166
13	257.4	95.0	535	14	CD707143	CD707143
14	221.6	81.8	522	12	BG44705	BG44705
15	221.6	81.8	653	10	BK637360	BK637360
16	221.6	81.8	1282	11	AK080144	AK080144
17	203	74.9	653	10	BF975096	BF975096
18	196	72.3	650	9	AJ449745	AJ449745
19	196	72.3	688	9	AJ450317	AJ450317
20	196	72.3	696	9	AJ453647	AJ453647
21	196	72.3	729	9	AJ450296	AJ450296
22	196	72.3	757	9	AJ446140	AJ446140
23	183.6	67.7	623	9	AJ450295	AJ450295
24	118.8	43.8	632	14	CD471262	CD471262
25	118.6	43.8	197	14	CD286686	CD286686
26	114	42.1	197	14	CF722800	CF722800
27	114	42.1	544	14	CD472096	CD472096
28	114	42.1	549	14	CD469966	CD469966
29	114	42.1	555	14	CD470596	CD470596
30	114	42.1	561	14	CD471717	CD471717
31	114	42.1	562	14	CD470721	CD470721
32	114	42.1	569	14	CD471750	CD471750
33	114	42.1	578	14	CD471811	CD471811
34	114	42.1	582	14	CD472071	CD472071
35	114	42.1	596	14	CD471678	CD471678
36	114	42.1	597	14	CD470940	CD470940
37	114	42.1	630	14	CD469721	CD469721
38	114	42.1	637	14	CD470714	CD470714
39	114	42.1	650	14	CD471162	CD471162
40	114	42.1	707	14	CD470713	CD470713
41	112.4	41.5	582	14	CD469582	CD469582
42	112.4	41.5	643	14	CD465467	CD465467
43	112.4	41.5	644	14	CD469636	CD469636
44	112.4	41.5	647	14	CD469225	CD469225
45	110.8	40.9	640	14	CD470661	CD470661

## ALIGNMENTS

RESULT 1  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

BF238155 541 bp mRNA linear EST 14-NOV-2000  
601811880F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4054915 5', mRNA sequence.  
BF238155  
BF238155.1 GI:11152074  
EST.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 541)  
NIH-MGC <http://mgs.nci.nih.gov/>  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Inceye Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Plate: LHC695 row: P column: 20  
High quality sequence stop: 541.



## FEATURES

source

Location/Qualifiers

1..541  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4054915"  
/tissue\_type="Primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_1ib="NIH MGC 48"  
/note="Organ: B-cells; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 100.0%; Score 271; DB 10; Length 541;  
Best Local Similarity 100.0%; Pred. No. 2.5e-47;  
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACGGTCGACAGTGAATTGCTTCCCGCTACATCTGGACTGGACCTAGACCT 60  
DB 234 AACGGTCGACAGTGAATTGCTTCCCGCTACATCTGGACTGGACCTAGACCT 293  
QY 61 GGCCTGCTACCGGCTACCTGTTCACTCTGGAGCCCTGCTACGACTGTGCCGA 120  
DB 294 GGCCTGCTACCGGCTACCTGTTCACTCTGGAGCCCTGCTACGACTGTGCCGA 353  
QY 121 CATGTGGCGGACTTCTGTGGAGGAAACCCCACTGAGGATCTTACCGCGCGC 180  
DB 354 CATGTGGCGGACTTCTGTGGAGGAAACCCCACTGAGGATCTTACCGCGCGC 413  
QY 181 CTCTACTTCTGTGAGGACCGCAAGGCTAGCCCGAGGGCTGCGCGCTACCGCGCC 240  
DB 414 CTCTACTTCTGTGAGGACCGCAAGGCTAGCCCGAGGGCTGCGCGCTACCGCGCC 473  
QY 241 GGGGTGCAATATGCAATCATGACCTTCAAG 271  
DB 474 GGGGTGCAATATGCAATCATGACCTTCAAG 504

RESULT 2  
BG757392 693 bp mRNA linear EST 15-MAY-2001  
LOCUS 602711022F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4851580 5',  
DEFINITION mRNA sequence.

ACCESSION BG757392  
VERSION BG757392.1 GI:14068045  
KEYWORDS EST.

SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
TITLE Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
JOURNAL NIH-MGC http://img.ncbi.nih.gov/  
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/HLN at:  
http://image.jnl.gov  
Plate: LNCM1634 row: k column: 05  
High quality sequence stop: 693.  
Location/Qualifiers

## FEATURES

## source

1..693  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4851580"  
/tissue\_type="Primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_1ib="NIH MGC 48"  
/note="Organ: B-cells; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 100.0%; Score 271; DB 12; Length 693;  
Best Local Similarity 100.0%; Pred. No. 2.7e-47;  
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACGGTCGACAGTGAATTGCTTCCCGCTACATCTGGAGCTGGACCTAGACCT 60  
DB 238 AACGGTCGACAGTGAATTGCTTCCCGCTACATCTGGAGCTGGACCTAGACCT 297  
QY 61 GGCCTGCTACCGGCTACCTGTTCACTCTGGAGCCCTGCTACGACTGTGCCGA 120  
DB 298 GGCCTGCTACCGGCTACCTGTTCACTCTGGAGCCCTGCTACGACTGTGCCGA 357  
QY 121 CATGTGGCGGACTTCTGTGGAGGAAACCCCACTGAGGATCTTACCGCGCGC 180  
DB 358 CATGTGGCGGACTTCTGTGGAGGAAACCCCACTGAGGATCTTACCGCGCGC 417  
QY 181 CTCTACTTCTGTGAGGACCGCAAGGCTAGCCCGAGGGCTGCGCGCTACCGCGCC 240  
DB 418 CTCTACTTCTGTGAGGACCGCAAGGCTAGCCCGAGGGCTGCGCGCTACCGCGCC 477  
QY 241 GGGGTGCAATATGCAATCATGACCTTCAAG 271  
DB 478 GGGGTGCAATATGCAATCATGACCTTCAAG 508

RESULT 3  
BG686133 743 bp mRNA linear EST 01-MAY-2001  
LOCUS 602638412F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:476234 5',  
DEFINITION mRNA sequence.

ACCESSION BG686133  
VERSION BG686133.1 GI:13917530  
KEYWORDS EST.

SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
TITLE Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
JOURNAL NIH-MGC http://img.ncbi.nih.gov/  
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/HLN at:  
http://image.jnl.gov  
Plate: LNCM1626 row: g column: 03  
High quality sequence stop: 740.  
Location/Qualifiers

## FEATURES

source 1..743



/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4766234"  
/issue\_type="Primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_11b="NIH\_MGC\_48"  
/note="Organ: B-cells; Vector: pOTB7; Site\_1: XhoI;  
Site\_2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCAACGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by Ling  
Hong in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 100.0%; Score 271; DB 12; Length 743;  
Best Local Similarity 100.0%; Pred.No. 2.8e-47;  
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACGGCTGCACGATGGAATTGCTCTCCCTACATCTCGAATCGGACCTAGACCT 60  
DB 234 AACGGCTGCACGATGGAATTGCTCTCCCTACATCTCGAATCGGACCTAGACCT 293  
QY 61 GGCCGCTGTACCGCGCTCACTGTTCACTCCTGGAGCCCTGTACGACTGTGCCGA 120  
DB 294 GGCCGCTGTACCGCGCTCACTGTTCACTCCTGGAGCCCTGTACGACTGTGCCGA 353  
QY 121 CATGTGGCCGACCTTTCTGGAGGAAACCCCACTCAGTGTAGAGATCTTCAACCGCGCC 180  
DB 354 CATGTGGCCGACCTTTCTGGAGGAAACCCCACTCAGTGTAGAGATCTTCAACCGCGCC 413  
QY 181 CTCTACTTCTGTAGAGACCGCAAGGCTGAGCGGCGCTGCGCGCTGACCGCGCC 240  
DB 414 CTCTACTTCTGTAGAGACCGCAAGGCTGAGCGGCGCTGCGCGCTGACCGCGCC 473  
QY 241 GGGGTGCAATAGCCATCATGACCTTCAAG 271  
DB 474 GGGGTGCAATAGCCATCATGACCTTCAAG 504

RESULT 4  
BX757089 820 bp mRNA linear EST 15-MAY-2001  
LOCUS 602715124FL NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4855517 5',  
DEFINITION mRNA sequence.  
ACCESSION BX757089  
VERSION BX757089.1 GI:14067742  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 820)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs@mail.nih.gov  
Tissue Procurement: Louis N. Staedt, M.D., Ph.D.  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
http://lncm1704 row: 0 column: 06  
High quality sequence scop: 675.  
Location/Qualifiers  
1..820  
/organism="Homo sapiens"

FEATURES  
source

/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4855517"  
/issue\_type="Primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_11b="NIH\_MGC\_48"  
/note="Organ: B-cells; Vector: pOTB7; Site\_1: XhoI;  
Site\_2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCAACGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by Ling  
Hong in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 100.0%; Score 271; DB 12; Length 820;  
Best Local Similarity 100.0%; Pred.No. 2.9e-47;  
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACGGCTGCACGATGGAATTGCTCTCCCTACATCTCGAATCGGACCTAGACCT 60  
DB 241 AACGGCTGCACGATGGAATTGCTCTCCCTACATCTCGAATCGGACCTAGACCT 300  
QY 61 GGCCGCTGTACCGCGCTCACTGTTCACTCCTGGAGCCCTGTACGACTGTGCCGA 120  
DB 301 GGCCGCTGTACCGCGCTCACTGTTCACTCCTGGAGCCCTGTACGACTGTGCCGA 360  
QY 121 CATGTGGCCGACCTTTCTGGAGGAAACCCCACTCAGTGTAGAGATCTTCAACCGCGCC 180  
DB 361 CATGTGGCCGACCTTTCTGGAGGAAACCCCACTCAGTGTAGAGATCTTCAACCGCGCC 420  
QY 181 CTCTACTTCTGTAGAGACCGCAAGGCTGAGCGGCGCTGCGCGCTGACCGCGCC 240  
DB 421 CTCTACTTCTGTAGAGACCGCAAGGCTGAGCGGCGCTGCGCGCTGACCGCGCC 480  
QY 241 GGGGTGCAATAGCCATCATGACCTTCAAG 271  
DB 481 GGGGTGCAATAGCCATCATGACCTTCAAG 511

RESULT 5  
BX464579 853 bp mRNA linear EST 22-MAY-2003  
LOCUS BX464579 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA  
DEFINITION clone CS0D0003YB14 5-PRIME, mRNA sequence.  
ACCESSION BX464579  
VERSION BX464579.1 GI:31031641  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 853)  
AUTHORS Li, M.B., Gruber, C., Jesse, J. and Polyes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 6672.r For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CSID0002F10Q1&cluster=6672.r. Contact :  
Feng Liang Email: fliang@lifetech.com URL:  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Paradise Avenue Genoscope sequence ID: CSID0002F10Q1.  
Location/Qualifiers  
1..853  
/organism="Homo sapiens"

FEATURES  
source



/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CSODG003YB14"  
/issue\_type="B CELLS (RAMOS CELL LINE)"  
/cell\_line="RAMOS CELL LINE"  
/clone\_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"  
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

## ORIGIN

Query Match 100.0%; Score 271; DB 13; Length 853;  
Best Local Similarity 100.0%; Pred. No. 2.9e-47;  
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACGGCTGCCAGTGAATTGCTCTTCTCCGCTACATCTCGAGCTGGAGCTAGACCT 60  
DB 242 AACGGCTGCCAGTGAATTGCTCTTCTCCGCTACATCTCGAGCTGGAGCTAGACCT 301  
QY 61 GGCCGCTGCTACCGGCTCACTGTTCACTCTCGAGCCCTGCTACGACTGTGCCGA 120  
DB 302 GGCCGCTGCTACCGGCTCACTGTTCACTCTCGAGCCCTGCTACGACTGTGCCGA 361  
QY 121 CATGTGCGCGACTTTCTGTGAGAGGAACCCCAACCTCAGTCTGAGATCTTGACGGCGGC 180  
DB 362 CATGTGCGCGACTTTCTGTGAGAGGAACCCCAACCTCAGTCTGAGATCTTGACGGCGGC 421  
QY 181 CTCTACTTCTGTGAGAGCGCAAGCTGAGCCCGAGGGCTGCGCGCTGCAACCGCGCC 240  
DB 422 CTCTACTTCTGTGAGAGCGCAAGCTGAGCCCGAGGGCTGCGCGCTGCAACCGCGCC 481  
QY 241 GGGGTGCAATAGCCATCATGACCTTCAAAG 271  
DB 482 GGGGTGCAATAGCCATCATGACCTTCAAAG 512

RESULT 6 872 bp mRNA linear EST 15-MAY-2001  
BG758510 60212721F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4851069 5',  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

Human sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
1 (bases 1 to 872)  
NIH-MGC http://mgi.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: LNCM1698 row: 1 column: 06  
High quality sequence stop: 836.  
Location/Qualifiers

## FEATURES

1..872  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4851069"  
/issue\_type="primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (phage-resistant)"

/clone\_lib="NIH\_MGC\_48"  
/note="Organ: B-cells; Vector: pOTB7; Site 1: XhoI;  
Site 2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Size-selected  
for average insert size 1.8kb. Library constructed by Ling  
Hong in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 100.0%; Score 271; DB 12; Length 872;  
Best Local Similarity 100.0%; Pred. No. 2.9e-47;  
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACGGCTGCCAGTGAATTGCTCTTCTCCGCTACATCTCGAGCTGGAGCTAGACCT 60  
DB 222 AACGGCTGCCAGTGAATTGCTCTTCTCCGCTACATCTCGAGCTGGAGCTAGACCT 281  
QY 61 GGCCGCTGCTACCGGCTCACTGTTCACTCTCGAGCCCTGCTACGACTGTGCCGA 120  
DB 282 GGCCGCTGCTACCGGCTCACTGTTCACTCTCGAGCCCTGCTACGACTGTGCCGA 341  
QY 121 CATGTGCGCGACTTTCTGTGAGAGGAACCCCAACCTCAGTCTGAGATCTTGACGGCGGC 180  
DB 342 CATGTGCGCGACTTTCTGTGAGAGGAACCCCAACCTCAGTCTGAGATCTTGACGGCGGC 401  
QY 181 CTCTACTTCTGTGAGAGCGCAAGCTGAGCCCGAGGGCTGCGCGCTGCAACCGCGCC 240  
DB 402 CTCTACTTCTGTGAGAGCGCAAGCTGAGCCCGAGGGCTGCGCGCTGCAACCGCGCC 461  
QY 241 GGGGTGCAATAGCCATCATGACCTTCAAAG 271  
DB 462 GGGGTGCAATAGCCATCATGACCTTCAAAG 492

RESULT 7 889 bp mRNA linear EST 01-MAY-2001  
BG686876 602650861F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4763247 5',  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

Human sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
1 (bases 1 to 889)  
NIH-MGC http://mgi.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: LNCM1618 row: 1 column: 16  
High quality sequence start: 6  
High quality sequence stop: 727.  
Location/Qualifiers

## FEATURES

1..889  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4763247"  
/issue\_type="primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (phage-resistant)"



/clone\_lib="NIH\_MGC\_48"  
/note="Organ: B-cells; Vector: pOTB7; Site\_1: XhoI;  
Site\_2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by Ling  
Hong in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 100.0%; Score 271; DB 12; Length 889;  
Best Local Similarity 100.0%; Pred. No. 2.9e-47;  
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACGGCTGCGACGTAATGCTCTTCCTCCGCTACATCTGGAGCTGAGACCTTACCT 60  
DB 194 AACGGCTGCGACGTAATGCTCTTCCTCCGCTACATCTGGAGCTGAGACCTTACCT 253  
QY 61 GGGCGGTGCTACCGGCTGACCTGCTTCACTCTGGAGCCCTGCTACGACTGTGCCGA 120  
DB 254 GGGCGGTGCTACCGGCTGACCTGCTTCACTCTGGAGCCCTGCTACGACTGTGCCGA 313  
QY 121 CATGTGGCCGACCTTTCTGCGAGGAAACCCCACTCAGTCTGAGATCTTCAACCGCGGC 180  
DB 314 CATGTGGCCGACCTTTCTGCGAGGAAACCCCACTCAGTCTGAGATCTTCAACCGCGGC 373  
QY 181 CTCTACTTCTGTAGAGACCGCAAGCTGAGCCGAGGCGGCTGACCGCGCC 240  
DB 374 CTCTACTTCTGTAGAGACCGCAAGCTGAGCCGAGGCGGCTGACCGCGCC 433  
QY 241 GGGGTGCAATAGCCATCATGACCTTCAAG 271  
DB 434 GGGGTGCAATAGCCATCATGACCTTCAAG 464

RESULT 8  
B0065440 953 bp mRNA linear EST 02-APR-2002  
LOCUS B0065440  
DEFINITION AGENCOURT 6855061 NIH\_MGC\_99 Homo sapiens cDNA clone IMAGE:5929977.  
5', mRNA sequence.  
ACCESSION B0065440  
VERSION B0065440.1 GI:19894486  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 953)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: Lou Staudt  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: LNCM2108 row: P column: 10  
High quality sequence stop: 634.  
Location/Qualifiers

FEATURES  
source

1. 953  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5929977"  
/tissue\_type="lymphoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_99"

/note="Organ: lymph; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCACAG(G). Size-selected >500bp for average insert size  
1.8kb. Library constructed by Ling Hong in the laboratory  
of Gerald M. Rubin (University of California, Berkeley)  
using ZAP-cDNA synthesis kit (Stratagene) and Superscript  
II RT (Life Technologies). Note: this is a NIH\_MGC  
Library."

## ORIGIN

Query Match 100.0%; Score 271; DB 13; Length 953;  
Best Local Similarity 100.0%; Pred. No. 3e-47;  
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACGGCTGCGACGTAATGCTCTTCCTCCGCTACATCTGGAGCTGAGACCTTACCT 60  
DB 220 AACGGCTGCGACGTAATGCTCTTCCTCCGCTACATCTGGAGCTGAGACCTTACCT 279  
QY 61 GGGCGGTGCTACCGGCTGACCTGCTTCACTCTGGAGCCCTGCTACGACTGTGCCGA 120  
DB 280 GGGCGGTGCTACCGGCTGACCTGCTTCACTCTGGAGCCCTGCTACGACTGTGCCGA 339  
QY 121 CATGTGGCCGACCTTTCTGCGAGGAAACCCCACTCAGTCTGAGATCTTCAACCGCGGC 180  
DB 340 CATGTGGCCGACCTTTCTGCGAGGAAACCCCACTCAGTCTGAGATCTTCAACCGCGGC 399  
QY 181 CTCTACTTCTGTAGAGACCGCAAGCTGAGCCGAGGCGGCTGACCGCGCC 240  
DB 400 CTCTACTTCTGTAGAGACCGCAAGCTGAGCCGAGGCGGCTGACCGCGCC 459  
QY 241 GGGGTGCAATAGCCATCATGACCTTCAAG 271  
DB 460 GGGGTGCAATAGCCATCATGACCTTCAAG 490

RESULT 9  
B0055935 1052 bp mRNA linear EST 29-MAR-2002  
LOCUS B0055935  
DEFINITION AGENCOURT\_6796291 NIH\_MGC\_99 Homo sapiens cDNA clone IMAGE:5808181  
5', mRNA sequence.  
ACCESSION B0055935  
VERSION B0055935.1 GI:19815262  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1052)  
AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: Lou Staudt  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: LNCM2051 row: M column: 14  
High quality sequence stop: 665.  
Location/Qualifiers

FEATURES  
source

1. 1052  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5808181"  
/tissue\_type="lymphoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_99"  
/note="Organ: lymph; Vector: pOTB7; Site\_1: XhoI; Site\_2:"



EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(G). Size-selected >500bp for average insert size 1.8Kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 100.0%; Score 271; DB 13; Length 1052;  
Best Local Similarity 100.0%; Pred. No. 3.1e-47;  
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACGGTGGCAGTGAATTGCTCTTCTCCGCTACATCTGGAGCTGGAGCTAGACCTT 60  
DB 220 AACGGTGGCAGTGAATTGCTCTTCTCCGCTACATCTGGAGCTGGAGCTAGACCTT 279  
QY 61 GGCGGTGTACCGGGTACCTGTTTCACTCTGAGAGCCCTGCTAGACCTGGCCGA 120  
DB 280 GGCGGTGTACCGGGTACCTGTTTCACTCTGAGAGCCCTGCTAGACCTGGCCGA 339  
QY 121 CATGTGGCCGACTTTCTGTGGAGGAAACCCCACTCACTGATGATCTTCAACCGCGCC 180  
DB 340 CATGTGGCCGACTTTCTGTGGAGGAAACCCCACTCACTGATGATCTTCAACCGCGCC 399  
QY 181 CTCTACTTCTGTGAGAGACCGCAAGGCTGAGCGGAGGCTGCGGCTGACCGCGCC 240  
DB 400 CTCTACTTCTGTGAGAGACCGCAAGGCTGAGCGGAGGCTGCGGCTGACCGCGCC 459  
QY 241 GGGGTGCAATAGCCATCATGACTTTCAAG 271  
DB 460 GGGGTGCAATAGCCATCATGACTTTCAAG 490

RESULT 10  
AL559877 1201 bp mRNA linear EST 31-MAY-2003

LOCUS AL559877 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA  
DEFINITION clone CS0D003YB14 5-PRIME, mRNA sequence.

ACCESSION AL559877  
VERSION AL559877.2 GI:31284408  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 1201)  
Li, W.B., Gruber, C., Jessee, J., and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
On Feb 15, 2001 this sequence version replaced gi:12505793.

CONTACT: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: segre@genoscope.cns.fr; Web: www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 6672.r For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0D003DA070P1&cluster=6672.r. Contact :  
Feng Liang Email: fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0D003DA070P1.

FEATURES  
source location/Qualifiers  
1..1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0D003YB14"  
/cissue\_type="B CELLS (RAMOS CELL LINE)"  
/cell\_line="RAMOS CELL LINE"  
/clone\_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"

/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

## ORIGIN

Query Match 100.0%; Score 271; DB 9; Length 1201;  
Best Local Similarity 100.0%; Pred. No. 3.2e-47;  
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACGGTGGCAGTGAATTGCTCTTCTCCGCTACATCTGGAGCTGGAGCTAGACCTT 60  
DB 248 AACGGTGGCAGTGAATTGCTCTTCTCCGCTACATCTGGAGCTGGAGCTAGACCTT 307  
QY 61 GGCGGTGTACCGGGTACCTGTTTCACTCTGAGAGCCCTGCTAGACCTGGCCGA 120  
DB 308 GGCGGTGTACCGGGTACCTGTTTCACTCTGAGAGCCCTGCTAGACCTGGCCGA 367  
QY 121 CATGTGGCCGACTTTCTGTGGAGGAAACCCCACTCACTGATGATCTTCAACCGCGCC 180  
DB 368 CATGTGGCCGACTTTCTGTGGAGGAAACCCCACTCACTGATGATCTTCAACCGCGCC 427  
QY 181 CTCTACTTCTGTGAGAGACCGCAAGGCTGAGCGGAGGCTGCGGCTGACCGCGCC 240  
DB 428 CTCTACTTCTGTGAGAGACCGCAAGGCTGAGCGGAGGCTGCGGCTGACCGCGCC 487  
QY 241 GGGGTGCAATAGCCATCATGACTTTCAAG 271  
DB 488 GGGGTGCAATAGCCATCATGACTTTCAAG 518

RESULT 11  
BX402063

LOCUS BX402063 1201 bp mRNA linear EST 13-MAY-2003  
DEFINITION BX402063 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED  
Homo sapiens cDNA clone CS0D012YD18 5-PRIME, mRNA sequence.

ACCESSION BX402063  
VERSION BX402063.1 GI:30626645  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 1201)  
Li, W.B., Gruber, C., Jessee, J., and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: segre@genoscope.cns.fr; Web: www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 6672.r For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0D012DB09P1&cluster=6672.r. Contact :  
Feng Liang Email: fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0D012DB09P1.

CONTACT: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: segre@genoscope.cns.fr; Web: www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 6672.r For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0D012DB09P1&cluster=6672.r. Contact :  
Feng Liang Email: fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0D012DB09P1.

FEATURES  
source location/Qualifiers  
1..1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0D012YD18"  
/cissue\_type="B CELLS (RAMOS CELL LINE)"  
/cell\_line="RAMOS CELL LINE"  
/clone\_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"  
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

FEATURES  
source location/Qualifiers  
1..1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0D012YD18"  
/cissue\_type="B CELLS (RAMOS CELL LINE)"  
/cell\_line="RAMOS CELL LINE"  
/clone\_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"  
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."



## ORIGIN

sites of the PCMVSPORT 6 vector. Library was normalized."

Query Match 100.0%; Score 271; DB 13; Length 1201;  
Best Local Similarity 100.0%; Pred. No. 3.2e-47;  
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACGGCTGCCACGTGGAATTGCTCTTCCCTCCGCTACATCTCGAAGCTGAGACCTT 60  
DB 308 AACGGCTGCCACGTGGAATTGCTCTTCCCTCCGCTACATCTCGAAGCTGAGACCTT 367  
QY 61 GGGCGCTGCTACCGCGCTCACTGCTTCACTCTGAGAGCCCTGCTACGACTGTGCCGA 120  
DB 368 GGGCGCTGCTACCGCGCTCACTGCTTCACTCTGAGAGCCCTGCTACGACTGTGCCGA 427  
QY 121 CATGTGGCCGACTTTCTGCGAGGGAACCCCACTGATGTGAGATCTTCCACCGCGCC 180  
DB 428 CATGTGGCCGACTTTCTGCGAGGGAACCCCACTGATGTGAGATCTTCCACCGCGCC 487  
QY 181 CTCTACTTCTGTGAGAGCCGCAAGCTGAGCCGAGGGCTGCGGCGCTGCAACGCGCC 240  
DB 488 CTCTACTTCTGTGAGAGCCGCAAGCTGAGCCGAGGGCTGCGGCGCTGCAACGCGCC 547  
QY 241 GGGGTGCAATATGACCATCATGACCTTCAAG 271  
DB 548 GGGGTGCAATATGACCATCATGACCTTCAAG 578

## RESULT 12

BF975166 942 bp mRNA linear EST 22-JAN-2001  
LOCUS 602244657F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4335639 5',  
DEFINITION mRNA sequence.

ACCSSION BF975166.1 GI:12342381  
VERSION BF975166  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
TITLE 1 (bases 1 to 942)  
JOURNAL NIH-MGC http://imgc.ncl.nih.gov/.  
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-remail.nih.gov  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LICM1207 row: a column: 16  
High quality sequence stop: 707.  
Location/Qualifiers

## FEATURES

source  
1. 942  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4335639"  
/tissue\_type="primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_id="NIH MGC 48"  
/note="Organ: B-cells; Vector: pORF7; Site: 1; XhoI;  
Site 2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCAGGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by Ling  
Hong in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 99.4%; Score 269.4; DB 10; Length 942;  
Best Local Similarity 99.6%; Pred. No. 6.5e-47;  
Matches 270; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AACGGCTGCCACGTGGAATTGCTCTTCCCTCCGCTACATCTCGAAGCTGAGACCTT 60  
DB 235 AACGGCTGCCACGTGGAATTGCTCTTCCCTCCGCTACATCTCGAAGCTGAGACCTT 294  
QY 61 GGGCGCTGCTACCGCGCTCACTGCTTCACTCTGAGAGCCCTGCTACGACTGTGCCGA 120  
DB 295 GGGCGCTGCTACCGCGCTCACTGCTTCACTCTGAGAGCCCTGCTACGACTGTGCCGA 354  
QY 121 CATGTGGCCGACTTTCTGCGAGGGAACCCCACTGATGTGAGATCTTCCACCGCGCC 180  
DB 355 CATGTGGCCGACTTTCTGCGAGGGAACCCCACTGATGTGAGATCTTCCACCGCGCC 414  
QY 181 CTCTACTTCTGTGAGAGCCGCAAGCTGAGCCGAGGGCTGCGGCGCTGCAACGCGCC 240  
DB 415 CTCTACTTCTGTGAGAGCCGCAAGCTGAGCCGAGGGCTGCGGCGCTGCAACGCGCC 474  
QY 241 GGGGTGCAATATGACCATCATGACCTTCAAG 271  
DB 475 GGGGTGCAATATGACCATCATGACCTTCAAG 505

## RESULT 13

CD707143 535 bp mRNA linear EST 25-JUN-2003  
LOCUS CD707143  
DEFINITION EST23670 human nasopharynx Homo sapiens cDNA, mRNA sequence.  
ACCSSION CD707143 GI:32237773  
VERSION CD707143.1  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
TITLE 1 (bases 1 to 535)  
JOURNAL Lin, X.-Q., Zhou, Y., Zhang, L.-J., Xu, H., Chen, H.-K., Pan, Z.-G. and  
Zeng, Y.-X.  
COMMENT Transcriptional Gene Expression Profile of Human Nasopharynx  
Unpublished (2003)  
Contact: Yixin Zeng  
Cancer Center  
Sun Yat-sen University  
651 Dongfeng Road East, Guangzhou 51060, China  
Tel: 86-1380-9770-743  
Fax: 86-20-8775-4506  
Email: yxzeng@gzsunu.edu.cn.  
Location/Qualifiers

## FEATURES

source  
1. 535  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/tissue\_type="normal nasopharynx"  
/clone\_id="human nasopharynx"  
/note="ESTs generated from a normal nasopharynx cDNA  
library from southern Chinese"

## ORIGIN

Query Match 95.0%; Score 257.4; DB 14; Length 535;  
Best Local Similarity 99.6%; Pred. No. 1.9e-44;  
Matches 258; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AACGGCTGCCACGTGGAATTGCTCTTCCCTCCGCTACATCTCGAAGCTGAGACCTT 60  
DB 277 AACGGCTGCCACGTGGAATTGCTCTTCCCTCCGCTACATCTCGAAGCTGAGACCTT 336  
QY 61 GGGCGCTGCTACCGCGCTCACTGCTTCACTCTGAGAGCCCTGCTACGACTGTGCCGA 120  
DB 337 GGGCGCTGCTACCGCGCTCACTGCTTCACTCTGAGAGCCCTGCTACGACTGTGCCGA 396



QY 121 CAGTGGCCGACCTTTCTGCGAGGAAACCCCACTGATGAGATCTTACCGCGCC 180  
 DB 397 CGGTGGCCGACCTTTCTGCGAGGAAACCCCACTGATGAGATCTTACCGCGCC 456  
 QY 181 CTCTACTCTGTGTAGGACCGCAAGCGTGAACCCGAGGGGTGTGGGGCTGACCGCGCC 240  
 DB 457 CTCTACTCTGTGTAGGACCGCAAGCGTGAACCCGAGGGGTGTGGGGCTGACCGCGCC 516  
 QY 241 GGGGTGCAATATGACATCA 259  
 DB 517 GGGGTGCAATATGACATCA 535

RESULT 14  
 BGI44705 522 bp mRNA linear EST 01-FEB-2001  
 LOCUS U73107.y1 Soares\_mouse\_MMB\_bcell Mus musculus cDNA clone  
 DEFINITION IMAGE:333637 5' similar to TR:09WVE0 Q9WVE0 ACTIVATION-INDUCED  
 CTTIDINE DEAMINASE, mRNA sequence.

ACCESSION BGI44705.1 GI:12648105  
 VERSION BGI44705  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 522)  
 NCICGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 JOURNAL Contact: Robert Strausberg, Ph.D.  
 COMMENT Email: cga@bcr-remail.nih.gov  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
 MGI:1077801  
 Seq primer: -40RP from Gibco  
 High quality sequence stop: 487.  
 Location/Qualifiers  
 1..522  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:333637"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="Soares\_mouse\_MMB\_bcell"  
 /note="Organ: germinal B-cell; Vector: pT7T3D-Pac  
 (Pharmacia) with a modified polylinker Site 1: Not I;  
 Site 2: Eco RI; 1st strand cDNA was primed with a Not I -  
 oligo(dT) primer [5'  
 TGTACCAATCTGAAGTGGAGGCGCGCTGTGTTTTTTTTTTTTTTTTTTTTT  
 T 3']; double-stranded cDNA was ligated to Eco RI  
 adaptors (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of the modified pT7T3D vector.  
 Library is normalized; constructed by Bento Soares and  
 M.Fatima Bonaudo."

ORIGIN  
 Query Match 81.8%; Score 221.6; DB 12; Length 522;  
 Best Local Similarity 89.2%; Pred. No. 7e-37;  
 Matches 239; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 184 TACTTCTGTAGGACCGCAAGGCTGAGCCCGAGGGGTGTGGGGCTGACCGCGCCGG 243  
 DB 257 TACTTCTGTAGGACCGCAAGGCTGAGCCCGAGGGGTGTGGGGCTGACCGCGCCGG 316  
 QY 244 GTCCAGATCGGATCATGACCTTCAAG 271  
 DB 317 GTCCAGATCGGATCATGACCTTCAAG 344

RESULT 15  
 B637360 653 bp mRNA linear EST 26-OCT-2001  
 LOCUS B637360  
 DEFINITION B637360 RIKEN full-length enriched, adult male aorta and vein Mus  
 musculus cDNA clone A530070C03 5', mRNA sequence.

ACCESSION B637360.1 GI:16473145  
 VERSION B637360  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 653)  
 Arai, K., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T.,  
 Hara, A., Hiramoto, K., Hori, F., Ishii, Y., Ito, M., Kawai, J.,  
 Kono, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K.,  
 Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K.,  
 Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
 Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,  
 Takeda, Y., Tanaka, T., Toyota, T., Muramatsu, M., and Hayashizaki, Y.  
 RIKEN Mouse ESTs (Arai, K., et al. 2001)  
 Unpublished (2001)

JOURNAL Contact: Yoshihide Hayashizaki  
 COMMENT Laboratory for Genome Exploration Research Group, RIKEN Genomic  
 Sciences Center (GSC), Yokohama Institute  
 The Institute of Physical and Chemical Research (RIKEN)  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 Tel: 81-45-503-9222  
 Fax: 81-45-503-9216  
 Email: genome-res@sc.riken.go.jp,  
 url:<http://genome.gsc.riken.go.jp/>,  
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
 Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
 Normalization and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new  
 genes. Genome Res. 10 (10), 1617-1630 (2000)  
 wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,  
 Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T.,  
 Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A.,  
 and Hayashizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system-384-format  
 sequencing pipeline with 384 multichannel sequencer. Genome Res.  
 10 (11), 1757-1771 (2000)  
 Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P.,  
 Sugahara, Y., and Hayashizaki, Y.  
 Computer-based methods for the mouse full-length cDNA  
 encyclopedia: real-time sequence clustering for construction of a  
 nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
 Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamahata, I.,  
 Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K., and  
 Hayashizaki, Y.  
 Computational Analysis of Full-Length Mouse cDNAs Compared with  
 Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)  
 Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
 further details.  
 e mouse tissues.  
 Location/Qualifiers  
 1..653  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="A530070C03"  
 /sex="male"

FEATURES  
 source



```

/tissue_type="aorta and vein"
/dev stage="adult"
/lab_host="DH10B"
/clone_lib="PIKEN full-length enriched, adult male aorta
and vein"

```

/note="Site 1: SalI; Site 2 BamHI. cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genetic Science Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer 5' GAGGAGAGAGAGATCCAGATCGTCTTTTTTTTTTTTTAA 3'. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 459.0. Second strand cDNA was prepared with the primer adapter of sequence 5' GAGGAGAGAGATTCGAGTTAAATTAATCCCCCCCCCCCC 3'. cDNA was cleaved with XhoI and BamHI. Vector: a modified Bluescript KS(+) after bulk excision from Lambda DCC I."

ORIGIN

Query Match	81.8%;	Score 221.6;	DB 10;	Length 653;
Best Local Similarity	89.2%;	Pred. No. 7.5e-37;		
Matches 239;	Conservative 0;	Mismatches 29;	Indels 0;	Gaps 0;

[illegible]

Search completed: March 13, 2004, 00:39:03  
Job time : 864.058 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: March 12, 2004, 12:27:59 ; Search time 365.199 Seconds  
(without alignments)  
13767.261 Million cell updates/sec

Title: US-09-966-880A-14

Sequence: 1 attatttttactgtctgcat.....agcttcggcgatcctcttctg 116

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

GenBank: 1: gb\_ba: 2: gb\_hcg: 3: gb\_in: 4: gb\_om: 5: gb\_ov: 6: gb\_pat: 7: gb\_ph: 8: gb\_pl: 9: gb\_pr: 10: gb\_ro: 11: gb\_sts: 12: gb\_sy: 13: gb\_un: 14: gb\_vl: 15: em\_ba: 16: em\_fun: 17: em\_hum: 18: em\_in: 19: em\_mu: 20: em\_om: 21: em\_or: 22: em\_ov: 23: em\_pat: 24: em\_ph: 25: em\_pl: 26: em\_ro: 27: em\_sts: 28: em\_un: 29: em\_vl: 30: em\_hcg\_hum: 31: em\_hcg\_inv: 32: em\_hcg\_other: 33: em\_hcg\_mus: 34: em\_hcg\_pln: 35: em\_hcg\_rtd: 36: em\_hcg\_man: 37: em\_hcg\_vtl: 38: em\_sy: 39: em\_hggo\_hum: 40: em\_hggo\_mus: 41: em\_hggo\_other:

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	* Query Match	Length	DB ID	Description
1	116	100.0	116	6 BD016839	BD016839 Novel cyt
2	116	100.0	335	6 HS577811	HS577811 Homo sapi
3	116	100.0	6564	6 BD016835	BD016835 Novel cyt
4	116	100.0	11204	6 BD016860	BD016860 Novel cyt
5	116	100.0	11204	6 AB040430	AB040430 Homo sapi
6	116	100.0	71132	9 AC092184	AC092184 Homo sapi
7	114.4	98.6	577	12 AF529840	AF529840 Mus muscu
8	114.4	98.6	591	12 AF529855	AF529855 Cricetulu
9	114.4	98.6	596	9 AF529816	AF529816 Homo sapi
10	114.4	98.6	596	9 AF529818	AF529818 Homo sapi
11	114.4	98.6	596	9 AF529819	AF529819 Homo sapi
12	114.4	98.6	596	9 AF529820	AF529820 Homo sapi
13	114.4	98.6	596	9 AF529821	AF529821 Homo sapi
14	114.4	98.6	596	9 AF529822	AF529822 Homo sapi
15	114.4	98.6	596	9 AF529823	AF529823 Homo sapi
16	114.4	98.6	596	9 AF529824	AF529824 Homo sapi
17	114.4	98.6	596	9 AF529825	AF529825 Homo sapi
18	114.4	98.6	596	9 AF529826	AF529826 Homo sapi
19	114.4	98.6	596	9 AF529827	AF529827 Homo sapi
20	114.4	98.6	596	12 AF529829	AF529829 Mus muscu
21	114.4	98.6	596	12 AF529830	AF529830 Mus muscu
22	114.4	98.6	596	12 AF529833	AF529833 Mus muscu
23	114.4	98.6	596	12 AF529835	AF529835 Mus muscu
24	114.4	98.6	596	12 AF529836	AF529836 Mus muscu
25	114.4	98.6	596	12 AF529837	AF529837 Mus muscu
26	114.4	98.6	596	12 AF529838	AF529838 Mus muscu
27	114.4	98.6	596	12 AF529839	AF529839 Mus muscu
28	114.4	98.6	597	12 BT007402	BT007402 Homo sapi
29	114.4	98.6	597	12 AF529828	AF529828 Mus muscu
30	114.4	98.6	597	12 AF529841	AF529841 Cricetulu
31	114.4	98.6	597	12 AF529842	AF529842 Cricetulu
32	114.4	98.6	597	12 AF529844	AF529844 Cricetulu
33	114.4	98.6	597	12 AF529845	AF529845 Cricetulu
34	114.4	98.6	597	12 AF529846	AF529846 Cricetulu
35	114.4	98.6	597	12 AF529847	AF529847 Cricetulu
36	114.4	98.6	597	12 AF529848	AF529848 Cricetulu
37	114.4	98.6	597	12 AF529849	AF529849 Cricetulu
38	114.4	98.6	597	12 AF529850	AF529850 Cricetulu
39	114.4	98.6	597	12 AF529852	AF529852 Cricetulu
40	114.4	98.6	597	12 AF529854	AF529854 Cricetulu
41	114.4	98.6	597	12 BT008226	BT008226 Synthetic
42	114.4	98.6	1828	9 BC006296	BC006296 Homo sapi
43	114.4	98.6	2791	9 AB040431	AB040431 Homo sapi
44	114.4	98.6	2818	6 BD016833	BD016833 Novel cyt
45	112.8	97.2	547	12 AF529856	AF529856 Cricetulu

## ALIGNMENTS

RESULT 1  
BD016839  
LOCUS BD016839 116 bp DNA linear PAT 27-AUG-2002  
DEFINITION Novel cytidine deaminase.  
ACCESSION BD016839  
VERSION BD016839.1 GI:22556015  
KEYWORDS UP 2001245669-A/12.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE  
AUTHORS Honjo, T. and Muramatsu, M.  
TITLE Novel cytidine deaminase  
JOURNAL Patent: JP 2001245669-A 12 11-SEP-2001;



COMMENT JAPAN TOBACCO INC. TASUKU HONJO  
OS Homo sapiens (human)  
PN JP 2001245669-A/12  
PD 11-SEP-2001  
PF 28-MAR-2000 JP 2000092981  
PI TASUKU HONJO, MASAMICHI MURAMATSU  
PC C12N15/09, A61K39/395, A61K39/395, A61P1/00, A61P11/06, A61P13/12,  
PC A61P17/00,  
PC A61P27/02, A61P27/16, A61P37/02, A61P37/08, C07K16/18, C12N1/19, PC  
C12N1/21,  
PC C12N5/10, C12N9/78, C12P21/02, C12P21/08, C12N1/21, C12R1/19, PC  
(C12N5/10, C12R1/91), C12N15/00, C12N5/00, C12N5/00, C12R1/91) CC  
FH Key Location/Qualifiers

FEATURES  
source  
1..116  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

ORIGIN  
Query Match 100.0%; Score 116; DB 6; Length 116;  
Best Local Similarity 100.0%; Pred. No. 2.6e-27;  
Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATATTTTCTGCTGGAATCTTTGTAGAAAACGAGAAAGACTTTCAGAGCCTGGG 60  
1 ATATTTTCTGCTGGAATCTTTGTAGAAAACGAGAAAGACTTTCAGAGCCTGGG 60

DB 1 ATATTTTCTGCTGGAATCTTTGTAGAAAACGAGAAAGACTTTCAGAGCCTGGG 60

QY 61 AAGGCTGCATGAAATTCAGTTCCTCCAGACAGCTTCGGCGCATCTTTTG 116  
61 AAGGCTGCATGAAATTCAGTTCCTCCAGACAGCTTCGGCGCATCTTTTG 116

Db 61 AAGGCTGCATGAAATTCAGTTCCTCCAGACAGCTTCGGCGCATCTTTTG 116

RESULT 2  
HSAS77811 335 bp mRNA linear prt 22-UTL-2003  
LOCUS Homo sapiens partial mRNA for activation-induced cytidine deaminase  
DEFINITION (Aid gene).  
ACCESSION AU577811.1 GI:33145978  
VERSION A577811.1  
KEYWORDS activation-induced cytidine deaminase; Aid gene.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1  
Takhar, P., Coker, H.A., Fear, D., Smurthwaite, L., Durham, S.R. and  
Gould, H.O.  
Allexgen and activation-induced deaminase drive heavy-chain class  
switching to IgH in the nasal mucosa of hay fever patients  
Unpublished  
2 (bases 1 to 335)  
Takhar, P.  
Direct Submission  
Submitted (21-UTL-2003) Takhar P., Biomedical Sciences Randall  
Centre, King's College London, New Hunt House, Guy's Campus, St  
Thomas St, London, SE1 1UL, UNITED KINGDOM  
Location/Qualifiers  
1..335  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/isolation\_source="nasal biopsy from inferior turbinate of  
an allergic rhinitis patient"  
/db\_xref="taxon:9606"

FEATURES  
source  
1..335  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

ORIGIN  
Query Match 100.0%; Score 116; DB 6; Length 116;  
Best Local Similarity 100.0%; Pred. No. 3e-27;  
Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATATTTTCTGCTGGAATCTTTGTAGAAAACGAGAAAGACTTTCAGAGCCTGGG 60  
1 ATATTTTCTGCTGGAATCTTTGTAGAAAACGAGAAAGACTTTCAGAGCCTGGG 60

DB 1 ATATTTTCTGCTGGAATCTTTGTAGAAAACGAGAAAGACTTTCAGAGCCTGGG 60

QY 61 AAGGCTGCATGAAATTCAGTTCCTCCAGACAGCTTCGGCGCATCTTTTG 116  
61 AAGGCTGCATGAAATTCAGTTCCTCCAGACAGCTTCGGCGCATCTTTTG 116

Db 61 AAGGCTGCATGAAATTCAGTTCCTCCAGACAGCTTCGGCGCATCTTTTG 116

RESULT 3  
BD016835 6564 bp DNA linear PAT 27-AUG-2002  
LOCUS Novel cytidine deaminase.  
DEFINITION BD016835  
ACCESSION BD016835.1 GI:22558011  
VERSION JP 2001245669-A/8.  
KEYWORDS Novel cytidine deaminase  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 6564)  
Honjo, T. and Muramatsu, M.  
Novel cytidine deaminase  
Patent: JP 2001245669-A 8 11-SEP-2001;  
JOURNAL JAPAN TOBACCO INC. TASUKU HONJO  
COMMENT OS Homo sapiens (human)  
PN JP 2001245669-A/8  
PD 11-SEP-2001  
PF 28-MAR-2000 JP 2000092981  
PI TASUKU HONJO, MASAMICHI MURAMATSU  
PC C12N15/09, A61K39/395, A61K39/395, A61P1/00, A61P11/06, A61P13/12,  
PC A61P17/00,  
PC A61P27/02, A61P27/16, A61P37/02, A61P37/08, C07K16/18, C12N1/19, PC  
C12N1/21,  
PC C12N5/10, C12N9/78, C12P21/02, C12P21/08, C12N1/21, C12R1/19, PC  
(C12N5/10, C12R1/91), C12N15/00, C12N5/00, C12N5/00, C12R1/91) CC  
FH Key Location/Qualifiers  
1..6564  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

ORIGIN  
Query Match 100.0%; Score 116; DB 6; Length 6564;  
Best Local Similarity 100.0%; Pred. No. 3e-27;  
Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATATTTTCTGCTGGAATCTTTGTAGAAAACGAGAAAGACTTTCAGAGCCTGGG 60  
1 ATATTTTCTGCTGGAATCTTTGTAGAAAACGAGAAAGACTTTCAGAGCCTGGG 60

DB 1 ATATTTTCTGCTGGAATCTTTGTAGAAAACGAGAAAGACTTTCAGAGCCTGGG 60

QY 61 AAGGCTGCATGAAATTCAGTTCCTCCAGACAGCTTCGGCGCATCTTTTG 116  
61 AAGGCTGCATGAAATTCAGTTCCTCCAGACAGCTTCGGCGCATCTTTTG 116

Db 3216 AAGGCTGCATGAAATTCAGTTCCTCCAGACAGCTTCGGCGCATCTTTTG 3216

/gene="AID"  
/function="editing (involved in somatic hypermutation and  
class switch recombination)"  
/codon\_start=3  
/evidence=experimental  
/product="activation-induced cytidine deaminase"  
/protein\_id="CA512261.1"  
/db\_xref="GI:33145979"  
/translation="DGRRCYRVWTFMSNCPYDCARVADFLRGNNLSLRIFETARY  
FEDRKAEPEGLRLIRAGVQVIAIMTFKDTFYCNMTFVNHETFRAMGLHNSVRL  
SRQLRLIL"

ORIGIN  
Query Match 100.0%; Score 116; DB 9; Length 335;  
Best Local Similarity 100.0%; Pred. No. 2.7e-27;  
Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATATTTTCTGCTGGAATCTTTGTAGAAAACGAGAAAGACTTTCAGAGCCTGGG 60  
1 ATATTTTCTGCTGGAATCTTTGTAGAAAACGAGAAAGACTTTCAGAGCCTGGG 60

DB 220 ATATTTTCTGCTGGAATCTTTGTAGAAAACGAGAAAGACTTTCAGAGCCTGGG 279

QY 61 AAGGCTGCATGAAATTCAGTTCCTCCAGACAGCTTCGGCGCATCTTTTG 116  
61 AAGGCTGCATGAAATTCAGTTCCTCCAGACAGCTTCGGCGCATCTTTTG 116

Db 280 AAGGCTGCATGAAATTCAGTTCCTCCAGACAGCTTCGGCGCATCTTTTG 335

RESULT 3  
BD016835 6564 bp DNA linear PAT 27-AUG-2002  
LOCUS Novel cytidine deaminase.  
DEFINITION BD016835  
ACCESSION BD016835.1 GI:22558011  
VERSION JP 2001245669-A/8.  
KEYWORDS Novel cytidine deaminase  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 6564)  
Honjo, T. and Muramatsu, M.  
Novel cytidine deaminase  
Patent: JP 2001245669-A 8 11-SEP-2001;  
JOURNAL JAPAN TOBACCO INC. TASUKU HONJO  
COMMENT OS Homo sapiens (human)  
PN JP 2001245669-A/8  
PD 11-SEP-2001  
PF 28-MAR-2000 JP 2000092981  
PI TASUKU HONJO, MASAMICHI MURAMATSU  
PC C12N15/09, A61K39/395, A61K39/395, A61P1/00, A61P11/06, A61P13/12,  
PC A61P17/00,  
PC A61P27/02, A61P27/16, A61P37/02, A61P37/08, C07K16/18, C12N1/19, PC  
C12N1/21,  
PC C12N5/10, C12N9/78, C12P21/02, C12P21/08, C12N1/21, C12R1/19, PC  
(C12N5/10, C12R1/91), C12N15/00, C12N5/00, C12N5/00, C12R1/91) CC  
FH Key Location/Qualifiers  
1..6564  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

ORIGIN  
Query Match 100.0%; Score 116; DB 6; Length 6564;  
Best Local Similarity 100.0%; Pred. No. 3e-27;  
Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATATTTTCTGCTGGAATCTTTGTAGAAAACGAGAAAGACTTTCAGAGCCTGGG 60  
1 ATATTTTCTGCTGGAATCTTTGTAGAAAACGAGAAAGACTTTCAGAGCCTGGG 60

DB 1 ATATTTTCTGCTGGAATCTTTGTAGAAAACGAGAAAGACTTTCAGAGCCTGGG 60

QY 61 AAGGCTGCATGAAATTCAGTTCCTCCAGACAGCTTCGGCGCATCTTTTG 116  
61 AAGGCTGCATGAAATTCAGTTCCTCCAGACAGCTTCGGCGCATCTTTTG 116

Db 3216 AAGGCTGCATGAAATTCAGTTCCTCCAGACAGCTTCGGCGCATCTTTTG 3216



RESULT 4  
BD016860  
LOCUS  
DEFINITION Novel cytidine deaminase.  
ACCESSION BD016860  
VERSION BD016860.1 GI:22558036  
KEYWORDS JP 2001245669-A/33.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Honjo, T. and Muramatsu, M.  
TITLE Novel cytidine deaminase.  
JOURNAL Patent: JP 2001245669-A 33 11-SEP-2001;  
JAPAN TOBACCO INC, TASUKU HONJO  
COMMENT  
OS Homo sapiens (human)  
PN JP 2001245669-A/33  
PD 11-SEP-2001  
PF 28-MAR-2000 JP 2000092981  
PI TASUKU HONJO, MASAMICHI MURAMATSU  
PC C12N15/09, A61K39/395, A61P1/00, A61P11/06, A61P13/12, A61P17/00,  
PC A61P27/02, A61P37/16, A61P37/02, A61P37/08, C07K16/18, C12N1/19, PC C12N1/21,  
PC C12N5/10, C12N9/78, C12P21/02, C12P21/08, C12N1/21, C12R1/19, PC C12N5/10, C12R1/91, C12N15/00, C12N5/00, C12N1/91 CC  
FH Key Location/Qualifiers  
FEATURES  
source 1..11204  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
ORIGIN  
Query Match 100.0%; Score 116; DB 6; Length 11204;  
Best Local Similarity 100.0%; Pred. No. 3.1e-27;  
Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATATTTTACTGCTGGAAATCTTTGTAGAAAACGAGAAAGACTTCAAGCTGGG 60  
DB 8371 ATATTTTACTGCTGGAAATCTTTGTAGAAAACGAGAAAGACTTCAAGCTGGG 8430  
QY 61 AAGGCTGCATGAATAATTCAGTTCCTCCAGACGCTTGGCGCATCCTTTG 116  
DB 8431 AAGGCTGCATGAATAATTCAGTTCCTCCAGACGCTTGGCGCATCCTTTG 8486  
RESULT 5  
AB040430  
LOCUS  
DEFINITION Homo sapiens A1D gene for activation-induced cytidine deaminase, complete cds.  
ACCESSION AB040430  
VERSION AB040430.1 GI:9988407  
KEYWORDS A1D; activation-induced cytidine deaminase.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Muto, T., Muramatsu, M., Taniwaki, M., Kinoshita, K. and Honjo, T.  
TITLE Isolation, tissue distribution, and chromosomal localization of the human activation-induced cytidine deaminase (A1D) gene  
JOURNAL Genomics 68 (1), 85-88 (2000)  
MEDLINE 20408890  
PUBMED 10950930  
REFERENCE  
AUTHORS Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanai, O., Catalan, N., Forveille, M., Dufourcq-Lagelouse, R., Gennery, A., Tzscan, I., Ersoy, F., Kayserili, H., Ugazio, A. G., Brousse, N., Muramatsu, M., Notarangelo, L. D., Kinoshita, K., Honjo, T., Fischer, A.

and Durandy, A.  
Activation-induced cytidine deaminase (A1D) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2)  
Cell 102 (5), 565-575 (2000)  
JOURNAL  
MEDLINE 20460541  
PUBMED 11007475  
REFERENCE  
AUTHORS Muto, T., Muramatsu, M., Taniwaki, M., Kinoshita, K. and Honjo, T.  
TITLE Direct Submission  
JOURNAL Submitted (18-MAR-2000) Tasuku Honjo, Kyoto University, Department of Medical Chemistry, Faculty of Medicine, Yoshida, Sakyo-ku, Kyoto, Kyoto 606-8501, Japan (E-mail: honjo@med.kyoto-u.ac.jp, Tel: 81-75-753-4371 (ex. 4371), Fax: 81-75-753-4388)  
FEATURES  
source 1..11204  
Location/Qualifiers  
1..11204  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/join(521..528, 6280..6427, 7807..8077, 8371..8486, 8956..9009)  
/gene="A1D"  
/join(521..528, 6280..6427, 7807..8077, 8371..8486, 8956..9009)  
/gene="A1D"  
/codon\_start=1  
/product="activation-induced cytidine deaminase"  
/protein\_id="BAB12720.1"  
/db\_xref="GI:9988408"  
/translation="MDSLMMRRKELVQFNVRWAKGRRETVLCYVKKRDSATPSLDFGLRNKGCHVELLFLRTISDDLDLPGRCYRTWTNTSPICDCAHVDPLRGNPMLSRLFTARLYFCEDRKAEPEGRLRHRRGVQALMTFDFYCWNTFVENHRTKAMGILHNSVRLSRQLRRILLPLYEVLDLDAFRTTL"  
ORIGIN  
Query Match 100.0%; Score 116; DB 2; Length 11204;  
Best Local Similarity 100.0%; Pred. No. 3.1e-27;  
Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATATTTTACTGCTGGAAATCTTTGTAGAAAACGAGAAAGACTTCAAGCTGGG 60  
DB 8371 ATATTTTACTGCTGGAAATCTTTGTAGAAAACGAGAAAGACTTCAAGCTGGG 8430  
QY 61 AAGGCTGCATGAATAATTCAGTTCCTCCAGACGCTTGGCGCATCCTTTG 116  
DB 8431 AAGGCTGCATGAATAATTCAGTTCCTCCAGACGCTTGGCGCATCCTTTG 8486  
RESULT 6  
AC092184  
LOCUS  
DEFINITION Homo sapiens 12 BAC RP11-438L7 (Roswell Park Cancer Institute Human BAC library) complete sequence.  
ACCESSION AC092184  
VERSION AC092184 AC013443  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C., Albrooks, S.L., Amarantunga, H.C., Are, J.R., Ayale, M., Banks, T., Barbarta, J., Benton, J., Birnagge, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowle, S., Brieve, M., Brown, E., Brown, M., Bryant, N.P., Bunay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Cartron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chiu, D., Chowdhury, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dahorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroil, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Emerling, S., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,



Hamilton, K., Han, J., Harris, C., Harris, K., Hart, M., Havlik, P., Haves, A., Hernandez, J., Hernandez, O., Hodgson, A., Hognes, M., Hollaway, C., Hollins, B., Homel, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Ioshikhes, I., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Jondah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Kovach, J., Kovar, C., Kratovic, J., Kureishi, A., Landry, N., Leal, B., Lee, E., Lewis, L.C., Lewis, L.J., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louised, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Maronde, I., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mel, G., Merscher, S., Metzger, M., Miller, A., Miner, G., Miner, Z., Mitchell, T., Mochabac, K., Montgomery, K.T., Morgan, N., Morris, S., Moser, M., Neal, D., Nelson, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okunnu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L., Quiles, M., Ren, Y., Rivers, M., Rojas, A., Rojudoan, I., Rolfe, M., Ruiz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shim, C., Shoshitari, N., Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabori, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansay, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vaequez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Kucherlapati, R., Weinstein, G., and Gibbs, R.

TITLE  
Direct Submission  
Unpublished  
2 (bases 1 to 71132)  
Worley, K.C.

REFERENCE  
AUTHORS  
JOURNAL  
Submitted (25-JUN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 71132)  
Worley, K.C.

REFERENCE  
AUTHORS  
JOURNAL  
Submitted (18-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
4 (bases 1 to 71132)  
Worley, K.C.

REFERENCE  
AUTHORS  
JOURNAL  
Submitted (25-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
5 (bases 1 to 71132)  
Worley, K.C.

REFERENCE  
AUTHORS  
JOURNAL  
Submitted (12-JUN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On May 25, 2002 this sequence version replaced gi:20901754.  
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email [gc-help@bcm.tmc.edu](mailto:gc-help@bcm.tmc.edu)

COMMENT

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

#### ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.  
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.  
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons

flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

#### FEATURES

source	QUALSTAT-REPORT
1..71132	Location/Qualifiers
/organism="Homo sapiens"	
/mol_type="genomic DNA"	
/db_xref="taxon:9606"	
/chromosome="12"	
/clone="RP11-438L7"	
/complement(1..1448)	
/note="overlaps bases 1..1448 of clone AC092490"	
/function="clone overlap"	
187..408	
/standard_name="57233"	
439..560	
/standard_name="92005"	
774..881	
/rpt_family="MIR"	
903..1190	
/rpt_family="AluSg"	
1191..1213	
/rpt_family="AT_rich"	
1591..1807	
/standard_name="6612"	
1744..1819	
/standard_name="8198"	
1966..2254	
/rpt_family="AluSg"	
3296..3328	
/rpt_family="TTTC)n"	
4097..4249	
/rpt_family="AluSg"	
4360..4654	
/rpt_family="AluY"	
/complement(5140..5262)	
/rpt_family="FLAM_C"	
5275..5304	
/rpt_family="AT_rich"	
5305..5563	
/rpt_family="AluYo"	
5564..5599	
/rpt_family="AT_rich"	
5644..5733	
/rpt_family="MSTD"	
5734..6012	
/rpt_family="AluSg"	
6020..6336	
/rpt_family="AluSx"	
6337..6629	
/rpt_family="MSTD"	
6630..6650	
/rpt_family="CAA)n"	
/complement(6711..6817)	
/rpt_family="MIR"	
/complement(7033..7327)	
/rpt_family="AluSx"	



```
repeat_region 7607..7745 /rpt_family="FLAM_A"
repeat_region 7980..8047 /rpt_family="MIR"
repeat_region complement(8050..8549) /rpt_family="LTR47A"
repeat_region 8567..8698 /rpt_family="MIR"
repeat_region complement(8815..9101) /rpt_family="AluSg"
repeat_region complement(11227..12175) /rpt_family="L1RS"
STS 12625..12900 /standard_name="87432"
repeat_region complement(13547..13664) /rpt_family="L1MB5"
repeat_region complement(13665..13979) /rpt_family="AluY"
repeat_region complement(13980..14193) /rpt_family="L1MB5"
repeat_region 14622..14924 /rpt_family="AluY"
repeat_region complement(15813..17136) /rpt_family="L1PAB"
repeat_region complement(17272..17395) /rpt_family="FLAM_C"
repeat_region 17426..17575 /rpt_family="MIR"
/rpt_family="MIR"
```

Query Match 100.0%; Score 116; DB 9; Length 71132;  
Best Local Similarity 100.0%; Pred. No. 3.3e-27;  
Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTATTTTACTGCTGGAATCTTTGTGAGAAACGACGAACTTTCAGGCTGG 60  
DB 42914 ATTATTTTACTGCTGGAATCTTTGTGAGAAACGACGAACTTTCAGGCTGG 42973  
DB 42974 AAGGCTGCATGAAATTCAGTCTCTCCAGACAGCTTCGGCGCATCTTTTG 43029

RESULT 7  
AF529840 577 bp mRNA linear SYN 17-SEP-2002  
LOCUS Mus musculus clone 13 transgenic Homo sapiens AID (AID) mRNA,  
DEFINITION complete cds.  
ACCESSION AF529840.1 GI:22297267  
VERSION AF529840.1 GI:22297267  
KEYWORDS Mus musculus (house mouse)  
SOURCE Mus musculus  
ORGANISM Mus musculus  
REFERENCE 1 (bases 1 to 577)  
AUTHORS Martin,A. and Scharff,M.D.  
TITLE Somatic hypermutation of the AID transgene in B and non-B cells  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12304-12308 (2002)  
REFERENCE 2 (bases 1 to 577)  
AUTHORS Martin,A. and Scharff,M.D.  
TITLE Direct Submision  
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of  
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA  
FEATURES  
source  
1..577  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="13"  
/cell\_type="hybridoma PI-5"  
/transgenic  
1..577  
/organism="Homo sapiens"

/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
<1..>577  
/gene="AID"  
<1..>577  
/gene="AID"  
/codon\_start=3  
/product="AID"  
/protein\_id="AA095427.1"  
/db\_xref="GI:22297268"  
/translation="RRKELYQFNVRWAKGRRENYLCYVKRRDSATSELDGRLN  
KNGCHVELPLRYSIDMDIDPGRCYRTWTSTSPCDCAH7ADPLKGNPNSLTF  
TARLYFCDERKAESEGRRLRAGVQALATFYQVTCMNTFVENHERITKAWEGJHE  
NSVRLSRQLRRILPLVEVDLDRDAFTTGL"

## ORIGIN

Query Match 98.6%; Score 114.4; DB 12; Length 577;  
Best Local Similarity 99.1%; Pred. No. 9.3e-27;  
Matches 115; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATTATTTTACTGCTGGAATCTTTGTGAGAAACGACGAACTTTCAGGCTGG 60  
DB 409 ATTATTTTACTGCTGGAATCTTTGTGAGAAACGACGAACTTTCAGGCTGG 468  
QY 61 AAGGCTGCATGAAATTCAGTCTCTCCAGACAGCTTCGGCGCATCTTTTG 116  
DB 469 AAGGCTGCATGAAATTCAGTCTCTCTCCAGACAGCTTCGGCGCATCTTTTG 524

RESULT 8  
AF529855 591 bp mRNA linear SYN 21-AUG-2003  
LOCUS Cricetus griseus transgenic clone 15 Homo sapiens AID (AID) mRNA,  
DEFINITION complete cds.  
ACCESSION AF529855  
VERSION AF529855.1 GI:22297297  
KEYWORDS Cricetus griseus (Chinese hamster)  
SOURCE Cricetus griseus  
ORGANISM Cricetus griseus  
REFERENCE 1 (bases 1 to 591)  
AUTHORS Martin,A. and Scharff,M.D.  
TITLE Somatic hypermutation of the AID transgene in B and non-B cells  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12304-12308 (2002)  
PUBMED 12202747  
REFERENCE 2 (bases 1 to 591)  
AUTHORS Martin,A. and Scharff,M.D.  
TITLE Direct Submision  
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of  
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA  
FEATURES  
source  
1..591  
/organism="Cricetus griseus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10029"  
/clone="15"  
/cell\_line="CHO"  
/transgenic  
1..591  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
<1..>591  
/gene="AID"  
/gene="AID"  
/codon\_start=1  
/product="AID"  
/protein\_id="AA095442.1"  
/db\_xref="GI:22297298"



ORIGIN

Query Match 98.6%; Score 114.4; DB 12; Length 591;  
Best Local Similarity 99.1%; Pred. No. 9.3e-27;  
Matches 115; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 ATATTCTTCTGCTGGAATCTTTGTAGAAAAACGACGAAGACTTCAAAAGCCTGGG 60  
422 ATATTCTTCTGCTGGAATCTTTGTAGAAAAACGACGAAGACTTCAAAAGCCTGGG 481

Db 61 AAGGCTGCATGAAATTCAGTCTCTCCAGACAGCTTCGGCGCATCTTTTG 116  
482 AAGGCTGCATGAAATTCAGTCTCTCCAGACAGCTTCGGCGCATCTTTTG 537

RESULT 9

AF529816 596 bp mRNA linear PRI 17-SEP-2002  
LOCUS Homo sapiens clone Ramos 2 AID (AID) mRNA, partial cds.  
AF529816  
VERSION AF529816.1 GI:22297219  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 596)  
TITLE Somatic hypermutation of the AID transgene in B and non-B cells  
AUTHORS Martin, A. and Scharff, M.D.  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12304-12308 (2002)  
REFERENCE 2 (bases 1 to 596)  
AUTHORS Martin, A. and Scharff, M.D.  
JOURNAL Direct Submission  
TITLE Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA  
FEATURES  
source  
1..596  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="Ramos 2"  
1..>596  
/gene="AID"  
1..>596  
/gene="AID"  
/note="Integrated into Burkitt's lymphoma cell line Ramos"  
/codon\_start=1  
/product="AID"  
/protein\_id="AA095403.1"  
/db\_xref="GI:22297220"  
/translation="MDSLMMRRKFLYQKRWAKGRETLYCYVVKRDSATSPSL  
DFGLRNKGCHEVLEFLRYISDMDDPGRCYVMTFTSRSPCYDCAHVDLRGNP  
NLSIRITATLAFCEDEKAPBEGRLRHRAGVOIAIMTFDYFCMNTVEVHERTFK  
AMEGLHNSVRLSRQLRILPLIEVDLRLDAFRITGL"

ORIGIN

Query Match 98.6%; Score 114.4; DB 9; Length 596;  
Best Local Similarity 99.1%; Pred. No. 9.3e-27;  
Matches 115; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 ATATTCTTCTGCTGGAATCTTTGTAGAAAAACGACGAAGACTTCAAAAGCCTGGG 60  
428 ATATTCTTCTGCTGGAATCTTTGTAGAAAAACGACGAAGACTTCAAAAGCCTGGG 487

Db 61 AAGGCTGCATGAAATTCAGTCTCTCCAGACAGCTTCGGCGCATCTTTTG 116  
488 AAGGCTGCATGAAATTCAGTCTCTCCAGACAGCTTCGGCGCATCTTTTG 543

ORIGIN

Query Match 98.6%; Score 114.4; DB 9; Length 596;  
Best Local Similarity 99.1%; Pred. No. 9.3e-27;  
Matches 115; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 ATATTCTTCTGCTGGAATCTTTGTAGAAAAACGACGAAGACTTCAAAAGCCTGGG 60  
428 ATATTCTTCTGCTGGAATCTTTGTAGAAAAACGACGAAGACTTCAAAAGCCTGGG 487

Db 61 AAGGCTGCATGAAATTCAGTCTCTCCAGACAGCTTCGGCGCATCTTTTG 116  
488 AAGGCTGCATGAAATTCAGTCTCTCCAGACAGCTTCGGCGCATCTTTTG 543

RESULT 10

AF529818 596 bp mRNA linear PRI 17-SEP-2002  
LOCUS Homo sapiens clone Ramos 4 truncated AID (AID) mRNA, complete cds.  
AF529818  
VERSION AF529818.1 GI:22297223  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 596)  
TITLE Somatic hypermutation of the AID transgene in B and non-B cells  
AUTHORS Martin, A. and Scharff, M.D.  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12304-12308 (2002)  
REFERENCE 2 (bases 1 to 596)  
AUTHORS Martin, A. and Scharff, M.D.  
JOURNAL Direct Submission  
TITLE Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA  
FEATURES  
source  
1..596  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="Ramos 4"  
1..596  
/gene="AID"  
1..570  
/gene="AID"  
/note="Integrated into Burkitt's lymphoma cell line Ramos"  
/codon\_start=1  
/product="truncated AID"  
/protein\_id="AA095405.1"  
/db\_xref="GI:22297224"  
/translation="MDSLMMRRKFLYQKRWAKGRETLYCYVVKRDSATSPSL  
DFGLRNKGCHEVLEFLRYISDMDDPGRCYVMTFTSRSPCYDCAHVDLRGNP  
NLSIRITATLAFCEDEKAPBEGRLRHRAGVOIAIMTFDYFCMNTVEVHERTFK  
AMEGLHNSVRLSRQLRILPLIEVDLRLDAFRITGL"

ORIGIN

Query Match 98.6%; Score 114.4; DB 9; Length 596;  
Best Local Similarity 99.1%; Pred. No. 9.3e-27;  
Matches 115; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 ATATTCTTCTGCTGGAATCTTTGTAGAAAAACGACGAAGACTTCAAAAGCCTGGG 60  
428 ATATTCTTCTGCTGGAATCTTTGTAGAAAAACGACGAAGACTTCAAAAGCCTGGG 487

Db 61 AAGGCTGCATGAAATTCAGTCTCTCCAGACAGCTTCGGCGCATCTTTTG 116  
488 AAGGCTGCATGAAATTCAGTCTCTCCAGACAGCTTCGGCGCATCTTTTG 543

ORIGIN

Query Match 98.6%; Score 114.4; DB 9; Length 596;  
Best Local Similarity 99.1%; Pred. No. 9.3e-27;  
Matches 115; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 ATATTCTTCTGCTGGAATCTTTGTAGAAAAACGACGAAGACTTCAAAAGCCTGGG 60  
428 ATATTCTTCTGCTGGAATCTTTGTAGAAAAACGACGAAGACTTCAAAAGCCTGGG 487

Db 61 AAGGCTGCATGAAATTCAGTCTCTCCAGACAGCTTCGGCGCATCTTTTG 116  
488 AAGGCTGCATGAAATTCAGTCTCTCCAGACAGCTTCGGCGCATCTTTTG 543

RESULT 11

AF529819 596 bp mRNA linear PRI 17-SEP-2002  
LOCUS Homo sapiens clone Ramos 5 AID (AID) mRNA, partial cds.  
AF529819  
VERSION AF529819.1 GI:22297225  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 596)  
TITLE Somatic hypermutation of the AID transgene in B and non-B cells  
AUTHORS Martin, A. and Scharff, M.D.  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12304-12308 (2002)  
REFERENCE 2 (bases 1 to 596)  
AUTHORS Martin, A. and Scharff, M.D.  
JOURNAL Direct Submission  
TITLE Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of



FEATURES  
Location/Qualifiers  
1..596

gene  
CDS  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="Ramos 5"  
1..>596  
/gene="AID"  
1..>596  
/note="Integrated into Burkitt's lymphoma cell line Ramos"  
/codon\_start=1  
/product="AID"  
/protein\_id="AA05406.1"  
/db\_xref="GI:22297226"  
/translation="MDSLNNRRKFLYQFNVRNAKGRRETYLCYVVKRDSATFSL  
DFGYLRNKGCHVELFLRYISDMDLPGRCYRTWTSPPCDCAHVAFLRGNP  
NLSIRIFPARLYFCVDRKABEGRLRLHRAVGQAIAIMTFKDYFCYCMNTEVENHERTEK  
AMEGHNENSVRLSRQLRIILPLIYVDLDRDAFRTIGL"

## ORIGIN

Query Match 98.6%; Score 114.4; DB 9; Length 596;  
Best Local Similarity 99.1%; Pred. No. 9.3e-27;  
Matches 115; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATTATTTTACTGCTGGAATACCTTTGAGAAAACCAAGAACTTTCAAGCTGGG 60  
DB 428 ATTATTTTACTGCTGGAATACCTTTGAGAAAACCAAGAACTTTCAAGCTGGG 487  
QY 61 AAGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGGCGCATCCTTTTG 116  
DB 488 AAGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGGCGCATCCTTTTG 543

RESULT 12  
LOCUS AF529820 596 bp mRNA linear PRI 17-SEP-2002  
DEFINITION Homo sapiens clone Ramos 6 AID (AID) mRNA, partial cds.  
ACCESSION AF529820  
VERSION AF529820.1 GI:22297227

## SOURCE

ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 596)  
AUTHORS Martin, A. and Schaff, M.D.  
TITLE Somatic hypermutation of the AID transgene in B and non-B cells  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12304-12308 (2002)  
AUTHORS 2 (bases 1 to 596)  
TITLE Direct Submission  
JOURNAL Martin, A. and Schaff, M.D.  
Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of  
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA

## FEATURES

source  
1..596  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="Ramos 6"  
1..>596  
/gene="AID"  
1..>596  
/note="Integrated into Burkitt's lymphoma cell line Ramos"  
/codon\_start=1  
/product="AID"  
/protein\_id="AA05407.1"  
/db\_xref="GI:22297228"

gene  
CDS  
/translation="MDSLNNRRKFLYQFNVRNAKGRRETYLCYVVKRDSATFSL  
DFGYLRNKGCHVELFLRYISDMDLPGRCYRTWTSPPCDCAHVAFLRGNP  
NLSIRIFPARLYFCVDRKABEGRLRLHRAVGQAIAIMTFKDYFCYCMNTEVENHERTEK  
AMEGHNENSVRLSRQLRIILPLIYVDLDRDAFRTIGL"

ORIGIN  
AMEGHNENSVRLSRQLRIILPLIYVDLDRDAFRTIGL"

Query Match 98.6%; Score 114.4; DB 9; Length 596;  
Best Local Similarity 99.1%; Pred. No. 9.3e-27;  
Matches 115; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATTATTTTACTGCTGGAATACCTTTGAGAAAACCAAGAACTTTCAAGCTGGG 60  
DB 428 ATTATTTTACTGCTGGAATACCTTTGAGAAAACCAAGAACTTTCAAGCTGGG 487  
QY 61 AAGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGGCGCATCCTTTTG 116  
DB 488 AAGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGGCGCATCCTTTTG 543

## RESULT 13

AF529821 596 bp mRNA linear PRI 17-SEP-2002  
LOCUS AF529821  
DEFINITION Homo sapiens clone Ramos 7 AID (AID) mRNA, partial cds.  
ACCESSION AF529821  
VERSION AF529821.1 GI:22297229

## SOURCE

ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 596)  
AUTHORS Martin, A. and Schaff, M.D.  
TITLE Somatic hypermutation of the AID transgene in B and non-B cells  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12304-12308 (2002)  
AUTHORS 2 (bases 1 to 596)  
TITLE Direct Submission  
JOURNAL Martin, A. and Schaff, M.D.  
Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of  
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA

## FEATURES

source  
1..596  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="Ramos 7"  
1..>596  
/gene="AID"  
1..>596  
/note="Integrated into Burkitt's lymphoma cell line Ramos"  
/codon\_start=1  
/product="AID"  
/protein\_id="AA05408.1"  
/db\_xref="GI:22297230"  
/translation="MDSLNNRRKFLYQFNVRNAKGRRETYLCYVVKRDSATFSL  
DFGYLRNKGCHVELFLRYISDMDLPGRCYRTWTSPPCDCAHVAFLRGNP  
NLSIRIFPARLYFCVDRKABEGRLRLHRAVGQAIAIMTFKDYFCYCMNTEVENHERTEK  
AMEGHNENSVRLSRQLRIILPLIYVDLDRDAFRTIGL"

## ORIGIN

Query Match 98.6%; Score 114.4; DB 9; Length 596;  
Best Local Similarity 99.1%; Pred. No. 9.3e-27;  
Matches 115; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATTATTTTACTGCTGGAATACCTTTGAGAAAACCAAGAACTTTCAAGCTGGG 60  
DB 428 ATTATTTTACTGCTGGAATACCTTTGAGAAAACCAAGAACTTTCAAGCTGGG 487  
QY 61 AAGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGGCGCATCCTTTTG 116  
DB 488 AAGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGGCGCATCCTTTTG 543

RESULT 14  
LOCUS AF529822 596 bp mRNA linear PRI 17-SEP-2002



DEFINITION Homo sapiens clone Ramos 8 AID (AID) mRNA, partial cds.  
ACCESSION AF529822 GI:22297231  
VERSION AF529822.1  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Martin, A. and Scharff, M.D.  
TITLE Somatic hypermutation of the AID transgene in B and non-B cells  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12304-12308 (2002)  
REFERENCE  
AUTHORS Martin, A. and Scharff, M.D.  
TITLE Direct Submision  
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of  
Medicine, 1300 Morris Park Ave. Bronx, NY 10461, USA  
FEATURES  
Location/Qualifiers  
1..596  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="Ramos 8"  
1..>596  
/gene="AID"  
1..>596  
/note="Integrated into Burkitt's lymphoma cell line Ramos"  
/codon\_start=1  
/product="AID"  
/protein\_id="AA095409.1"  
/db\_xref="GI:22297232"  
/translation="MDSLMNRKFLYQFNVMKGRRETYLCYVVKRDSATSFSL  
DFGLRNKGCHEVLEFLRYISDMDDPGRCYVWTFSWSPCYDCARVADFLRGNP  
NLSLRIFTLRYPCEDRKAPESGLRLHRAVGQIALMTFDYFCNNTFVENHERTFK  
AMEGLHNSVRLSRQLRIILPLYEVDDLRDAFRTIGL"  
ORIGIN  
Query Match 98.6%; Score 114.4; DB 9; Length 596;  
Best Local Similarity 99.1%; Pred. No. 9.3e-27;  
Matches 115; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 ATTATTTTACTGCTGGAATACCTTTGTAGAAAAACGAAAGACTTTCAAGCCTGGG 60  
DB 428 ATTATTTTACTGCTGGAATACCTTTGTAGAAAAACGAAAGACTTTCAAGCCTGGG 487  
QY 61 AAGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGGCGCATCCTTTTG 116  
DB 488 AAGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGGCGCATCCTTTTG 543  
RESULT 15  
AF529823 596 bp mRNA linear PRI 17-SEP-2002  
LOCUS Homo sapiens clone Ramos 9 AID (AID) mRNA, partial cds.  
DEFINITION AF529823  
ACCESSION AF529823 GI:22297233  
VERSION AF529823.1  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
AUTHORS Martin, A. and Scharff, M.D.  
TITLE Somatic hypermutation of the AID transgene in B and non-B cells  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12304-12308 (2002)  
REFERENCE  
AUTHORS Martin, A. and Scharff, M.D.  
TITLE Direct Submision  
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of  
Medicine, 1300 Morris Park Ave. Bronx, NY 10461, USA  
FEATURES  
Location/Qualifiers  
1..596

/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="Ramos 9"  
1..>596  
/gene="AID"  
1..>596  
/note="Integrated into Burkitt's lymphoma cell line Ramos"  
/codon\_start=1  
/product="AID"  
/protein\_id="AA095410.1"  
/db\_xref="GI:22297234"  
/translation="MDSLMNRKFLYQFNVMKGRRETYLCYVVKRDSATSFSL  
DFGLRNKGCHEVLEFLRYISDMDDPGRCYVWTFSWSPCYDCARVADFLRGNP  
NLSLRIFTLRYPCEDRKAPESGLRLHRAVGQIALMTFDYFCNNTFVENHERTFK  
AMEGLHNSVRLSRQLRIILPLYEVDDLRDAFRTIGL"  
ORIGIN

Query Match 98.6%; Score 114.4; DB 9; Length 596;  
Best Local Similarity 99.1%; Pred. No. 9.3e-27;  
Matches 115; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 ATTATTTTACTGCTGGAATACCTTTGTAGAAAAACGAAAGACTTTCAAGCCTGGG 60  
DB 428 ATTATTTTACTGCTGGAATACCTTTGTAGAAAAACGAAAGACTTTCAAGCCTGGG 487  
QY 61 AAGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGGCGCATCCTTTTG 116  
DB 488 AAGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGGCGCATCCTTTTG 543

Search completed: March 12, 2004, 21:22:31  
Job time : 365.199 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 12, 2004, 12:26:39 ; Search time 46.4502 Seconds

(without alignments)  
10609.014 Million cell updates/sec

Title: US-09-966-880A-14

Perfect score: 116  
Sequence: 1 attatcttaccgtcgtgaat.....agcttcggccatccttttg 116

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_29cand4:\*

1: geneseqn1980s:\*\n2: geneseqn1390s:\*\n3: geneseqn2000s:\*\n4: geneseqn2001as:\*\n5: geneseqn2001bs:\*\n6: geneseqn2002as:\*\n7: geneseqn2003as:\*\n8: geneseqn2003bs:\*\n9: geneseqn2003cs:\*\n10: geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	116	100.0	116	3 AAC55318	AAC55318 Human act
2	116	100.0	6564	3 AAC55314	AAC55314 Human act
3	116	100.0	11204	3 AAC55339	AAC55339 Human act
4	116	100.0	11204	6 AB873286	AB873286 DNA encod
5	114.4	98.6	557	8 ADB88952	ADB88952 A1D gene
6	114.4	98.6	1543	7 ABX05468	ABX05468 Human nov
7	114.4	98.6	2791	6 AB873287	AB873287 DNA encod
8	114.4	98.6	2791	6 AB873288	AB873288 DNA encod
9	114.4	98.6	2818	3 AAC55312	AAC55312 Human act
10	96.8	83.4	2440	3 AAC55307	AAC55307 Mouse act
11	34.2	29.5	1348	6 ABN96785	ABN96785 Gene #328
12	34.2	29.5	1446	2 AAA12409	AAA12409 CDNA enco
13	34.2	29.5	1534	2 AA220856	AA220856 Polynucle
14	34.2	29.5	1534	4 AA559293	AA559293 Human CDN
15	34.2	29.5	1534	6 ABA90962	ABA90962 Human pol
16	34.2	29.5	1777	6 ADD18987	ADD18987 Human dls
17	34.2	29.5	201143	6 ABK83568	ABK83568 Human DNA
18	30.4	26.2	823	9 ADE25658	ADE25658 Human CDN
19	30.4	26.2	1055	6 AAD24392	AAD24392 Human RNA
20	30.4	26.2	1427	9 AAK73067	AAK73067 Human imm
21	30.4	26.2	1526	9 ADD18895	ADD18895 Human dls
22	30.4	26.2	1567	4 AAI93080	AAI93080 Human pol
23	30.4	26.2	1638	6 ABQ61187	ABQ61187 Unidentif

24	30.2	26.0	945	6 ABQ69695	ABQ69695 Listeria
25	30.2	26.0	945	6 ABQ69758	ABQ69758 Listeria
26	30.2	26.0	966	6 ABQ67740	ABQ67740 Listeria
27	30.2	26.0	966	6 ABQ67741	ABQ67741 Listeria
28	30.2	26.0	84707	6 ABQ67196_6	ABQ67196_6
29	30.2	26.0	110000	6 ABQ67196_1	ABQ67196_1
30	30.2	26.0	110000	6 ABQ69245_24	ABQ69245_24
31	30.2	26.0	110000	6 ABQ69245_26	ABQ69245_26
32	29	25.0	7113	4 ABA09050	ABA09050
33	29	25.0	7113	4 AAK52465	AAK52465
34	29	25.0	7287	4 AAK51481	AAK51481
35	28.8	24.8	274	5 ABV89929	ABV89929 Human pro
36	28.6	24.7	15181	4 AAK65588	AAK65588 Human imm
37	28.6	24.7	15185	4 AAK65589	AAK65589 Human imm
38	28.6	24.7	68230	8 ADA66349	ADA66349 Mouse ppp
39	28.6	24.7	68233	8 ADA03065	ADA03065 Mouse ppp
40	28.6	24.7	68233	9 ADB72803	ADB72803 Mouse ppp
41	28.2	24.3	743	4 AAI94932	AAI94932 Human neu
42	28	24.1	28	3 AAC55332	AAC55332 Human act
43	28	24.1	2000	7 ADA71431	ADA71431 Rice gene
44	27.8	24.0	261	4 AAS24582	AAS24582 Human ova
45	27.8	24.0	390	5 AAH83200	AAH83200 Human ova

## ALIGNMENTS

RESULT 1  
AAC55318  
ID AAC55318 standard; DNA, 116 BP.

AC AAC55318;

DT 05-FEB-2001 (first entry)

DE Human activation-induced cytidine deaminase exon 4 SEQ ID NO:14.

KM Activation-induced cytidine deaminase; A1D; cytidine deaminase;  
KM Immune related disease; allergy; allergic disease; antiallergic;  
KM antianemic; antiaesthetic; ophthalmological; anti-HIV; dermatological;  
KM gene therapy; B cell associated immune system disorder; food allergy;  
KM immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
KM IgA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;  
KM drug allergy; allergic rhinitis; Rosen disease; Pigeon disease; AIDS;  
KM ataxia telangiectasia; common variable immunodeficiency disorder;  
KM major histocompatibility class II deficiency disease;  
KM auto immunodeficiency syndrome; IgG subclass selection disorder; ds.

OS Homo sapiens.

PN WO200058480-A1.

PD 05-OCT-2000.

PF 28-MAR-2000; 2000WO-JP001918.

PR 29-MAR-1999; 99JP-00087192.

PR 24-JUN-1999; 99JP-00178999.

PR 27-DEC-1999; 99JP-00371382.

PA (NTSB) JAPAN TOBACCO INC.

PI (HONJ) HONJO T.

DR Honjo T, Muramatsu M;

PT Nucleic acid encoding activation induced cytidine deaminase, useful as a  
PT target for drug development for immune-related diseases including  
PT allergies.

FS Claim 18; Page 151; 174pp; Japanese.

XX



CC The present invention describes an activation-induced cytidine deaminase  
CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and has  
CC cytidine activity similar to APOBEC-1. AID has anti-allergic, anti-naemic,  
CC antiasthmatic, ophthalmological, anti-HIV and dermatological activities,  
CC and can be used in gene therapy. AID polynucleotides are useful in  
CC methods for identifying drugs for the treatment of B cell associated  
CC immune system disorders, immunodeficiency diseases and allergies, such as  
CC immunoglobulin A (IGA) deficiency disease, IGA nephritis, gamma-  
CC globulinaemia, atopic dermatitis, allergic colitis, asthma, food allergy,  
CC drug allergy, allergic rhinitis, Rosen disease, Digorge disease, ataxia  
CC telangiectasia, common variable immunodeficiency disorder, MHC (major  
CC histocompatibility class) class II deficiency disease, AIDS (auto  
CC immunodeficiency syndrome), elevated Igg disorder, and Igg subclass  
CC selection disorder. The DNA sequences encoding AID may be used for gene  
CC therapy and the antibodies to the AID protein may be used for diagnosis  
CC and treatment of these disorders. The present sequence represents the  
CC exon 4 genomic DNA sequence of human AID  
XX

SQ Sequence 116 BP; 32 A; 25 C; 24 G; 35 T; 0 U; 0 Other;

Query Match 100.0%; Score 116; DB 3; Length 116;

Best Local Similarity 100.0%; Pred. No. 3.6e-29; Mismatches 0; Gaps 0;

Matches 116; Conservative 0; Indels 0; Gaps 0;

1 ATTATTTTACTGCTGGAATCTTTGTAGAAAACACGAAAGAACTTCAAGCTGGG 60

1 ATTATTTTACTGCTGGAATCTTTGTAGAAAACACGAAAGAACTTCAAGCTGGG 60

61 AAGGGCTGCATGAAATTCAGTTCGTCTCCAGACAGCTTGGGGCATCTTTTG 116

61 AAGGGCTGCATGAAATTCAGTTCGTCTCCAGACAGCTTGGGGCATCTTTTG 116

61 AAGGGCTGCATGAAATTCAGTTCGTCTCCAGACAGCTTGGGGCATCTTTTG 116

RESULT 2  
AACS5314  
ID AACS5314 standard; DNA; 6564 BP.

XX AACS5314;

AC AACS5314;

DE 05-FEB-2001 (first entry)

XX Human activation-induced cytidine deaminase genomic DNA SEQ ID NO:10.

XX Activation-induced cytidine deaminase; AID; cytidine deaminase;

XX immune related disease; allergy; allergic disease; antiallergic;

XX antianaemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;

XX gene therapy; B cell associated immune system disorder; food allergy;

XX immunodeficiency disease; immunoglobulin A deficiency disease; asthma;

XX IGA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;

XX drug allergy; allergic rhinitis; Rosen disease; Digorge disease; AIDS;

XX ataxia telangiectasia; common variable immunodeficiency disorder;

XX major histocompatibility class II deficiency disease;

XX auto immunodeficiency syndrome; Igg subclass selection disorder; ds.

OS Homo sapiens.

PN WO200058480-A1.

PD 05-OCT-2000.

PF 28-MAR-2000; 2000WO-JP001918.

XX 29-MAR-1999; 99JP-00087192.

XX 24-JUN-1999; 99JP-00178999.

XX 27-DEC-1999; 99JP-00371382.

XX (NISR ) JAPAN TOBACCO INC.

XX (HONJ/) HONJO T.

XX Honjo T, Muramatsu M;

XX WPI, 2000-611715/58.

PT Nucleic acid encoding activation induced cytidine deaminase, useful as a  
PT target for drug development for immune-related diseases including  
PT allergies.

XX Claim 17; Page 145-150; 174pp; Japanese.

XX The present invention describes an activation-induced cytidine deaminase  
XX (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and has  
XX cytidine activity similar to APOBEC-1. AID has anti-allergic, anti-naemic,  
XX antiasthmatic, ophthalmological, anti-HIV and dermatological activities,  
XX and can be used in gene therapy. AID polynucleotides are useful in  
XX methods for identifying drugs for the treatment of B cell associated  
XX immune system disorders, immunodeficiency diseases and allergies, such as  
XX immunoglobulin A (IGA) deficiency disease, IGA nephritis, gamma-  
XX globulinaemia, atopic dermatitis, allergic colitis, asthma, food allergy,  
XX drug allergy, allergic rhinitis, Rosen disease, Digorge disease, ataxia  
XX telangiectasia, common variable immunodeficiency disorder, MHC (major  
XX histocompatibility class) class II deficiency disease, AIDS (auto  
XX immunodeficiency syndrome), elevated Igg disorder, and Igg subclass  
XX selection disorder. The DNA sequences encoding AID may be used for gene  
XX therapy and the antibodies to the AID protein may be used for diagnosis  
XX and treatment of these disorders. The present sequence represents a  
XX genomic DNA sequence of human AID  
XX

SQ Sequence 6564 BP; 1909 A; 1358 C; 1383 G; 1914 T; 0 U; 0 Other;

Query Match 100.0%; Score 116; DB 3; Length 6564;

Best Local Similarity 100.0%; Pred. No. 1.3e-28; Mismatches 0; Gaps 0;

Matches 116; Conservative 0; Indels 0; Gaps 0;

1 ATTATTTTACTGCTGGAATCTTTGTAGAAAACACGAAAGAACTTCAAGCTGGG 60

1 ATTATTTTACTGCTGGAATCTTTGTAGAAAACACGAAAGAACTTCAAGCTGGG 60

3156 ATTATTTTACTGCTGGAATCTTTGTAGAAAACACGAAAGAACTTCAAGCTGGG 3215

61 AAGGGCTGCATGAAATTCAGTTCGTCTCCAGACAGCTTGGGGCATCTTTTG 116

61 AAGGGCTGCATGAAATTCAGTTCGTCTCCAGACAGCTTGGGGCATCTTTTG 116

RESULT 3  
AACS5339  
ID AACS5339 standard; DNA; 11204 BP.

XX AACS5339;

AC AACS5339;

DE 05-FEB-2001 (first entry)

XX Human activation-induced cytidine deaminase genomic DNA SEQ ID NO:35.

XX Activation-induced cytidine deaminase; AID; cytidine deaminase;

XX immune related disease; allergy; allergic disease; antiallergic;

XX antianaemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;

XX gene therapy; B cell associated immune system disorder; food allergy;

XX immunodeficiency disease; immunoglobulin A deficiency disease; asthma;

XX IGA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;

XX drug allergy; allergic rhinitis; Rosen disease; Digorge disease; AIDS;

XX ataxia telangiectasia; common variable immunodeficiency disorder;

XX major histocompatibility class II deficiency disease;

XX auto immunodeficiency syndrome; Igg subclass selection disorder; ds.

OS Homo sapiens.

PN WO200058480-A1.

PD 05-OCT-2000.

PF 28-MAR-2000; 2000WO-JP001918.

XX 29-MAR-1999; 99JP-00087192.

XX 24-JUN-1999; 99JP-00178999.

XX 27-DEC-1999; 99JP-00371382.

XX (NISR ) JAPAN TOBACCO INC.



PA (HONU/) HONTO T.  
 XX  
 XX Honjo T, Muramatsu M;  
 XX  
 DR WPI; 2000-611715/58.  
 XX  
 PT Nucleic acid encoding activation induced cytidine deaminase, useful as a  
 PT target for drug development for immune-related diseases including  
 PT allergies.  
 XX  
 PS Claim 17; Page 163-170; 174pp; Japanese.  
 XX  
 CC The present invention describes an activation-induced cytidine deaminase  
 CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and has  
 CC cytidine activity similar to APOBEC-1. AID has antiallergic, antianemic,  
 CC antiasthmatic, ophthalmological, anti-HIV and dermatological activities,  
 CC and can be used in gene therapy. AID polynucleotides are useful in  
 CC methods for identifying drugs for the treatment of B cell associated  
 CC immune system disorders, immunodeficiency diseases and allergies, such as  
 CC immunoglobulin A (IgA) deficiency disease, IgA nephritis, gamma-  
 CC globulinemia, atopic dermatitis, allergic colitis, asthma, food allergy,  
 CC drug allergy, allergic rhinitis, Rosen disease, DisGeorge disease, ataxia  
 CC telangiectasia, common variable immunodeficiency disorder, MHC (major  
 CC histocompatibility class) class II deficiency disease, AIDS (auto  
 CC immunodeficiency syndrome), elevated IgE disorder, and IgG subclass  
 CC selection disorder. The DNA sequences encoding AID may be used for gene  
 CC therapy and the antibodies to the AID protein may be used for diagnosis  
 CC and treatment of these disorders. The present sequence represents a  
 CC genomic DNA sequence of human AID  
 XX  
 SQ Sequence 11204 BP; 3305 A; 2273 C; 2373 G; 3253 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 116; DB 3; Length 11204;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-28;  
 Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATTATTTTACTGCTGGAATCTTTGTAGAAAACCAAGAAAGACTTTCAAGCTGGG 60  
 DB 8371 ATTATTTTACTGCTGGAATCTTTGTAGAAAACCAAGAAAGACTTTCAAGCTGGG 8430  
 QY 61 AAGGCTGCATGAAATTCAGTTCCTCTCCAGACAGCTTCGGCGATCCTTTTG 116  
 DB 8431 AAGGCTGCATGAAATTCAGTTCCTCTCCAGACAGCTTCGGCGATCCTTTTG 8486  
 RESULT 4  
 ABS73286  
 ID ABS73286 standard; DNA; 11204 BP.  
 XX  
 AC ABS73286;  
 XX  
 DT 04-DEC-2002 (first entry)  
 XX  
 DE DNA encoding human translocation del(12p) protein #1.  
 XX  
 KW Chromosome aberration; oncogenic fusion protein; cancer;  
 KW proliferative disease; cellular protein isoform; heat shock protein 90;  
 KW HSP-90; rheumatoid arthritis; cancer; haematopoietic disorder;  
 KW T cell lymphoma; B cell lymphoma; chronic myeloid leukaemia; CML;  
 KW acute lymphoid leukaemia; ALL; chronic myelomonocytic leukaemia; CMML;  
 KW acute lymphoblastic leukaemia; ALL; APL; NHL; solid tumour;  
 KW papillary thyroid carcinoma; Ewing's sarcoma; melanoma; liposarcoma;  
 KW rhabdomyosarcoma; synovial sarcoma; viral infection; gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 EN WO200269900-A2.  
 XX  
 PD 12-SEP-2002.  
 XX  
 PF 01-MAR-2002; 2002WO-US006518.  
 XX  
 PR 01-MAR-2001; 2001US-0272751P.

XX  
 PA (CONF-) CONFORMA THERAPEUTICS CORP.  
 XX  
 XX Fritz LC, Burrows FJ;  
 PI  
 XX  
 DR WPI; 2002-698710/75.  
 XX  
 DR F-PSDB; ABG95082.  
 XX  
 PT Treating genetically-defined disease associated with chromosomal  
 PT aberrations yielding oncogenic fusion proteins, e.g. cell proliferative  
 PT diseases, involves administering an inhibitor of heat shock protein 90.  
 XX  
 PS Disclosure; Page 242-245; 389pp; English.  
 XX  
 CC The invention describes a method of treating genetically-defined disease  
 CC associated with chromosomal aberrations yielding oncogenic fusion  
 CC proteins (I), treating cancerous cells containing (I) in a heterogeneous  
 CC cell population, treating proliferative diseases associated with mutant  
 CC protein or cellular protein isoforms (II) dependent on heat shock protein  
 CC (HSP)-90, or selectively treating cells expressing (II) involving  
 CC administering HSP90-inhibitor. The method is useful for treating  
 CC genetically-defined disease with chromosomal aberration yielding  
 CC oncogenic fusion protein, treating cancerous cells containing fusion  
 CC protein in heterogeneous cell population, treating proliferative disease  
 CC (e.g. rheumatoid arthritis or cancer) associated with mutant protein or  
 CC cellular protein isoform dependent on heat shock protein (HSP)-90 (e.g.  
 CC p53), or selectively treating cells expressing mutant protein or cellular  
 CC protein isoform in a patient heterozygous for (II). The method is useful  
 CC for treating a disease e.g. haematopoietic disorder such as T or B cell  
 CC lymphoma, chronic myeloid leukaemia (CML), APL, ALL, NHL and CML,  
 CC or a disease characterised by a solid tumour such as papillary thyroid  
 CC carcinoma, Ewing's sarcoma, melanoma, liposarcoma, rhabdomyosarcoma and  
 CC synovial sarcoma. The method is also useful for treating viral  
 CC infections. This represents the DNA sequence of a chromosome aberration  
 XX  
 SQ Sequence 11204 BP; 3305 A; 2273 C; 2373 G; 3253 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 116; DB 6; Length 11204;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-28;  
 Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATTATTTTACTGCTGGAATCTTTGTAGAAAACCAAGAAAGACTTTCAAGCTGGG 60  
 DB 8371 ATTATTTTACTGCTGGAATCTTTGTAGAAAACCAAGAAAGACTTTCAAGCTGGG 8430  
 QY 61 AAGGCTGCATGAAATTCAGTTCCTCTCCAGACAGCTTCGGCGATCCTTTTG 116  
 DB 8431 AAGGCTGCATGAAATTCAGTTCCTCTCCAGACAGCTTCGGCGATCCTTTTG 8486  
 RESULT 5  
 ADB88952  
 ID ADB88952 standard; DNA; 597 BP.  
 XX  
 AC ADB88952;  
 XX  
 DT 04-DEC-2003 (first entry)  
 XX  
 DE AID gene as substrate used in AID mutagenic method.  
 XX  
 KW ds; cytostatic; virucide; antidote; mutation;  
 KW activation-induced cytidine deaminase; cell phenotype;  
 KW monoclonal antibody; antigen; cross-reactivity; tumour;  
 KW germ warfare agent; toxin; V region.  
 XX  
 OS Homo sapiens.  
 XX  
 EN Key  
 XX  
 FT mutation  
 FT  
 FT mutation  
 FT  
 FT Location/Qualifiers  
 replace(9..A)  
 /\*tag= a  
 /\*note= "mutation in Pl-5 hybridoma cell"  
 replace(12..A)  
 /\*tag= b



```

FT mutation /note= "mutation in Ramos cell"
FT replace(14,C)
FT /*tag= c
FT /note= "mutation in CHO cell"
FT mutation replace(61,T)
FT /*tag= d
FT /note= "mutation in P1-5 hybridoma cell"
FT mutation replace(66,A)
FT /*tag= e
FT /note= "mutation in P1-5 hybridoma cell"
FT mutation replace(73,T)
FT /*tag= f
FT /note= "mutation in CHO cell"
FT mutation replace(79,G)
FT /*tag= g
FT /note= "mutation in Ramos cell"
FT mutation replace(85,G)
FT /*tag= h
FT /note= "mutation in P1-5 hybridoma cell"
FT mutation replace(112,C)
FT /*tag= i
FT /note= "mutation in Ramos cell"
FT mutation replace(118,G)
FT /*tag= j
FT /note= "mutation in P1-5 hybridoma cell"
FT mutation replace(156,T)
FT /*tag= k
FT /note= "mutation in P1-5 hybridoma cell"
FT mutation replace(161,A)
FT /*tag= l
FT /note= "mutation in P1-5 hybridoma cell"
FT mutation replace(165,T,G)
FT /*tag= m
FT /note= "G mutation in Ramos cell, T mutation in CHO cell"
FT mutation replace(178,T)
FT /*tag= n
FT /note= "mutation in Ramos cell"
FT mutation replace(206,G)
FT /*tag= o
FT /note= "mutation in CHO cell"
FT mutation replace(209,C)
FT /*tag= p
FT /note= "mutation in Ramos cell"
FT mutation replace(218,T)
FT /*tag= q
FT /note= "mutation in Ramos cell"
FT mutation replace(250,C)
FT /*tag= r
FT /note= "mutation in Ramos cell"
FT mutation replace(261,G)
FT /*tag= s
FT /note= "mutation in Ramos cell"
FT mutation replace(289,A)
FT /*tag= t
FT /note= "mutation in Ramos cell"
FT mutation replace(303,T)
FT /*tag= u
FT /note= "mutation in Ramos cell"
FT mutation replace(329,T)
FT /*tag= v
FT /note= "mutation in Ramos cell"
FT mutation replace(350,G,T)
FT /*tag= w
FT /note= "T mutation in Ramos cell, G mutation in CHO cell"
FT mutation replace(356,A)
FT /*tag= x
FT /note= "A mutation in Ramos cell, A mutation in CHO cell"
FT mutation replace(357,G)
FT /*tag= y
FT /note= "mutation in Ramos cell"
FT mutation replace(367,T, A)
FT /*tag= z
FT /note= "mutation in P1-5 hybridoma cell"

FT mutation replace(375,A)
FT /*tag= aa
FT /note= "mutation in CHO cell"
FT mutation replace(402,A)
FT /*tag= ab
FT /note= "mutation in P1-5 hybridoma cell"
FT mutation replace(410,T)
FT /*tag= ac
FT /note= "mutation in P1-5 hybridoma cell"
FT mutation replace(418,T)
FT /*tag= ad
FT /note= "mutation in P1-5 hybridoma cell"
FT mutation replace(420,T)
FT /*tag= ae
FT /note= "mutation in P1-5 hybridoma cell"
FT mutation replace(421,A)
FT /*tag= af
FT /note= "mutation in CHO cell"
FT mutation replace(430,G)
FT /*tag= ag
FT /note= "mutation in Ramos cell"
FT mutation replace(440,A)
FT /*tag= ah
FT /note= "mutation in CHO cell"
FT mutation replace(449,T)
FT /*tag= ai
FT /note= "mutation in P1-5 hybridoma cell"
FT mutation replace(495,A)
FT /*tag= aj
FT /note= "mutation in CHO cell"
FT mutation replace(517,C)
FT /*tag= ak
FT /note= "mutation in Ramos cell"
FT mutation replace(526,T, A)
FT /*tag= al
FT /note= "T mutation in P1-5 hybridoma cell, A mutation in Ramos cell"
FT mutation replace(533,A)
FT /*tag= am
FT /note= "mutation in P1-5 hybridoma cell"
FT mutation replace(538,A)
FT /*tag= an
FT /note= "mutation in CHO cell"
FT mutation replace(549,A)
FT /*tag= ao
FT /note= "mutation in CHO cell"
FT mutation replace(568,T)
FT /*tag= ap
FT /note= "mutation in Ramos cell"
FT mutation replace(570,T)
FT /*tag= aq
FT /note= "mutation in CHO cell"

W02003061363-A2.
31-JUL-2003.
15-JAN-2003; 2003WC-US001149.
17-JAN-2002; 2002US-0350269P.
(YESH ) UNIV YESHIVA EINSTEIN COLLEGE.
Martin A, Scharff MD;
WPI, 2003-747983/70.
Induction of mutations in gene expressed in eukaryotic cell useful e.g.
to produce antibodies with greater affinity or specificity for antigens
e.g. human pathogens by expressing an activation-induced cytidine
deaminase gene in cell.
Example 4, Fig 6; 73pp; English.

```



XX The invention relates to the induction of mutations in a gene expressed  
 CC in a eukaryotic cell, where the gene is operably linked to a promoter and  
 CC within 2 kb of promoter, by expressing a transgenic activation-induced  
 CC cytidine deaminase (AID) gene in the cell. The method is useful for  
 CC producing mutated genes and proteins, determining the effect of a  
 CC mutation on a protein or cell phenotype, and producing genes and proteins  
 CC with altered properties. It is especially useful for producing antibodies  
 CC with altered affinities or specificities for an antigen, or  
 CC increased/decreased cross-reactivity for a second antigen, e.g. to  
 CC produce antibodies with greater affinity/specificity useful  
 CC therapeutically to treat tumours, viruses such as Ebola and Lassa Fever  
 CC or against germ warfare agents, toxins (e.g. ricin). The method may  
 CC similarly be used to alter affinity, specificity or cross-reactivity of a  
 CC monoclonal antibody. This sequence represents the AID gene used in the  
 CC method of the invention to induce mutations in the sequence.

XX  
 SQ Sequence 597 BP; 128 A; 164 C; 155 G; 150 T; 0 U; 0 Other;

Query Match 98.6%; Score 114.4; DB 8; Length 597;  
 Best Local Similarity 99.1%; Pred. No. 2.1e-28;  
 Matches 115; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ATATATTTTACTGCTGGAATACCTTTGTAGAAAACCGAAGAACTTCAAGCTGGG 60  
 DB 428 ATTATTTTACTGCTGGAATACCTTTGTAGAAAACCGAAGAACTTCAAGCTGGG 487  
 OY 61 AAGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGGCGCATCTTTTG 116  
 DB 488 AAGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGGCGCATCTTTTG 543

RESULT 6  
 ABX05468  
 ID ABX05468 standard; cDNA; 1543 BP.

XX ABX05468;

XX 17-JAN-2003 (first entry)

XX Human novel polynucleotide #483.

XX Human; gene; ss; genetic disorder; gene mapping; medical imaging; cancer;  
 XX neurodegenerative disorder; lymphoid cell disorder; osteoporosis;  
 XX Parkinson's disease; Alzheimer's disease; bone degenerative disorder;  
 XX osteoarthritis; periodontal disease; liver fibrosis; viral infection;  
 XX fungal infection; bacterial infection; autoimmune disease; diabetes;  
 XX atopic dermatitis.

XX Homo sapiens.

XX WO200274961-A1.

XX 26-SEP-2002.

XX 14-MAR-2002; 2002WO-US005109.

XX 15-MAR-2001; 2001US-00810173.

XX (HYPE-) HYSEQ INC.

XX Tang YT, Zhou F, Goodrich R, Asundi V, Zhang J, Zhao QA, Ren F,  
 XX Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;  
 XX Wehrman T, Wang J, Wang D, Drmanac RT;

XX WPI; 2003-040556/03.

XX P-PSDB; ABU00390.

XX New isolated polypeptides and polynucleotides, useful for preventing,  
 XX treating or ameliorating medical conditions, such as cancer,  
 XX neurodegenerative disorders, lymphoid cell disorders, bone degenerative  
 XX disorders, and infections.

PS Claim 1; SEQ ID NO 483; 235pp; English.

XX The invention relates to human polynucleotides and the polypeptides they  
 CC encode. The polynucleotides and polypeptides are useful in diagnostics,  
 CC forensics, gene mapping, medical imaging, identification of mutations,  
 CC responsible for genetic disorders or other traits, assessing biodiversity  
 CC and producing many other types of data and products dependent on DNA and  
 CC amino acid sequences. They are also useful for preventing, treating or  
 CC ameliorating medical conditions, such as cancer, neurodegenerative  
 CC disorders (e.g. Parkinson's disease, Alzheimer's disease), lymphoid cell  
 CC disorders, osteoporosis, osteoarthritis, bone degenerative disorders,  
 CC periodontal disease, liver fibrosis, infections (e.g. viral, fungal or  
 CC bacterial) or autoimmune diseases (e.g. diabetes, atopic dermatitis).  
 CC Sequences ABX04986-ABX05511 represent human polynucleotides of the  
 CC invention. Note: The sequence data for this patent is not represented in  
 CC the printed specification but is based on sequence information supplied  
 CC by the European Patent Office

XX  
 SQ Sequence 1543 BP; 428 A; 334 C; 352 G; 429 T; 0 U; 0 Other;

Query Match 98.6%; Score 114.4; DB 7; Length 1543;  
 Best Local Similarity 99.1%; Pred. No. 2.8e-28;  
 Matches 115; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ATATATTTTACTGCTGGAATACCTTTGTAGAAAACCGAAGAACTTCAAGCTGGG 60  
 DB 505 ATTATTTTACTGCTGGAATACCTTTGTAGAAAACCGAAGAACTTCAAGCTGGG 564  
 OY 61 AAGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGGCGCATCTTTTG 116  
 DB 565 AAGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGGCGCATCTTTTG 620

RESULT 7  
 ABS73287  
 ID ABS73287 standard; DNA; 2791 BP.

XX ABS73287;

XX 04-DEC-2002 (first entry)

XX DNA encoding human translocation del(12p) protein #2.

XX Chromosome aberration; oncogenic fusion protein; cancer;  
 XX proliferative disease; cellular protein isoform; heat shock protein 90;  
 XX HSP-90; rheumatoid arthritis; cancer; haematopoietic disorder;  
 XX T cell lymphoma; B cell lymphoma; chronic myeloid leukaemia; CML;  
 XX acute myeloid leukaemia; AML; chronic myelomonocytic leukaemia; CMML;  
 XX acute lymphoblastic leukaemia; ALL; APL; NHL; solid tumour;  
 XX papillary thyroid carcinoma; Ewing's sarcoma; melanoma; liposarcoma;  
 XX rhabdomyosarcoma; synovial sarcoma; viral infection; gene; ds.

XX Homo sapiens.

XX WO200269900-A2.

XX 12-SEP-2002.

XX 01-MAR-2002; 2002WO-US006518.

XX 01-MAR-2001; 2001US-0272751P.

XX (CONF-) CONFORMA THERAPEUTICS CORP.

XX Fritz LC, Burrows FJ;

XX WPI; 2002-698710/75.

XX P-PSDB; ABG95083.

XX Treating genetically-defined disease associated with chromosomal  
 XX aberrations yielding oncogenic fusion proteins, e.g. cell proliferative  
 XX diseases, involves administering an inhibitor of heat shock protein 90.







PD 05-OCT-2000.  
 XX  
 PF 28-MAR-2000; 2000WO-JP001918.  
 XX  
 PR 29-MAR-1999; 99JP-00087192.  
 PR 24-JUN-1999; 99JP-00178999.  
 PR 27-DEC-1999; 99JP-00371382.  
 XX  
 PA (NIBS ) JAPAN TOBACCO INC.  
 PA (HONJ/) HONJO T.  
 XX  
 PI Honjo T, Muramatsu M;  
 XX  
 DR WPI; 2000-611715/58.  
 DR P-PSDB; AAB24198.  
 XX  
 PT Nucleic acid encoding activation induced cytidine deaminase, useful as a  
 PT target for drug development for immune-related diseases including  
 PT allergies.  
 XX  
 PS Claim 3; Page 135-139; 174pp; Japanese.  
 XX  
 CC The present sequence encodes human activation-induced cytidine deaminase  
 CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and has  
 CC cytidine activity similar to APOBEC-1. AID has anti-allergic, antianaemic,  
 CC antiasthmatic, ophthalmological, anti-HIV and dermatological activities,  
 CC and can be used in gene therapy. AID polynucleotides are useful in  
 CC methods for identifying drugs for the treatment of B cell associated  
 CC immune system disorders, immunodeficiency diseases and allergies, such as  
 CC immunoglobulin A (IGA) deficiency disease, IGA nephritis, gamma-  
 CC globulinaemia, atopic dermatitis, allergic colitis, asthma, food allergy,  
 CC drug allergy, allergic rhinitis, Rosen disease, Digorge disease, ataxia  
 CC telangiectasia, common variable immunodeficiency disorder, MHC (major  
 CC histocompatibility class II deficiency disease, AIDS (auto  
 CC immunodeficiency syndrome), elevated Ige disorder, and Ige subclass  
 CC selection disorder. The DNA sequences encoding AID may be used for gene  
 CC therapy and the antibodies to the AID protein may be used for diagnosis  
 CC and treatment of these disorders  
 XX  
 SQ Sequence 2818 BP; 868 A; 548 C; 626 G; 776 T; 0 U; 0 Other;  
 Query Match 98.6%; Score 114.4; DB 3; Length 2818;  
 Best Local Similarity 99.1%; Pred. No. 3.4e-28;  
 Matches 115; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 ATTAATTTTACTGCTGGAATACCTTTGTAGAAAACACGAAAGAACTTTCAAGCCTGGG 60  
 DB 507 ATTAATTTTACTGCTGGAATACCTTTGTAGAAAACACGAAAGAACTTTCAAGCCTGGG 566  
 QY 61 AAGGCGCTGATGAATAATTCAGTTGCTCTCTCCAGACAGCTTCGGCGATCCTTTTG 116  
 DB 567 AAGGCGCTGATGAATAATTCAGTTGCTCTCTCCAGACAGCTTCGGCGATCCTTTTG 622  
 XX  
 RESULT 10  
 AAC55307  
 ID AAC55307 standard; cDNA; 2440 BP.  
 XX  
 AC AAC55307;  
 XX  
 DT 05-FEB-2001 (first entry)  
 XX  
 DE Mouse activation-induced cytidine deaminase encoding cDNA SEQ ID NO:1.  
 XX  
 KW Activation-induced cytidine deaminase; AID; cytidine deaminase;  
 KW immune related disease; allergy; allergic disease; antiallergic;  
 KW antianaemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;  
 KW gene therapy; B cell associated immune system disorder; food allergy;  
 KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
 KW IGA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;  
 KW drug allergy; allergic rhinitis; Rosen disease; Digorge disease; AIDS;  
 KW ataxia telangiectasia; common variable immunodeficiency disorder;  
 KW major histocompatibility class II deficiency disease;

KW auto immunodeficiency syndrome; Ige subclass selection disorder; ss.  
 XX  
 OS Mus musculus.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 93..689  
 FT /\*tag= a  
 FT /product= "activation-induced cytidine deaminase"  
 XX  
 PN WO200058480-A1.  
 XX  
 PD 05-OCT-2000.  
 XX  
 PF 28-MAR-2000; 2000WO-JP001918.  
 XX  
 PR 29-MAR-1999; 99JP-00087192.  
 PR 24-JUN-1999; 99JP-00178999.  
 PR 27-DEC-1999; 99JP-00371382.  
 XX  
 PA (NIBS ) JAPAN TOBACCO INC.  
 PA (HONJ/) HONJO T.  
 XX  
 PI Honjo T, Muramatsu M;  
 XX  
 DR WPI; 2000-611715/58.  
 DR P-PSDB; AAB24197.  
 XX  
 PT Nucleic acid encoding activation induced cytidine deaminase, useful as a  
 PT target for drug development for immune-related diseases including  
 PT allergies.  
 XX  
 PS Claim 3; Page 126-130; 174pp; Japanese.  
 XX  
 CC The present sequence encodes mouse activation-induced cytidine deaminase  
 CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and has  
 CC cytidine activity similar to APOBEC-1. AID has anti-allergic, antianaemic,  
 CC antiasthmatic, ophthalmological, anti-HIV and dermatological activities,  
 CC and can be used in gene therapy. AID polynucleotides are useful in  
 CC methods for identifying drugs for the treatment of B cell associated  
 CC immune system disorders, immunodeficiency diseases and allergies, such as  
 CC immunoglobulin A (IGA) deficiency disease, IGA nephritis, gamma-  
 CC globulinaemia, atopic dermatitis, allergic colitis, asthma, food allergy,  
 CC drug allergy, allergic rhinitis, Rosen disease, Digorge disease, ataxia  
 CC telangiectasia, common variable immunodeficiency disorder, MHC (major  
 CC histocompatibility class II deficiency disease, AIDS (auto  
 CC immunodeficiency syndrome), elevated Ige disorder, and Ige subclass  
 CC selection disorder. The DNA sequences encoding AID may be used for gene  
 CC therapy and the antibodies to the AID protein may be used for diagnosis  
 CC and treatment of these disorders  
 XX  
 SQ Sequence 2440 BP; 706 A; 546 C; 551 G; 636 T; 0 U; 1 Other;  
 Query Match 83.4%; Score 96.8; DB 3; Length 2440;  
 Best Local Similarity 89.7%; Pred. No. 2.7e-22;  
 Matches 104; Conservative 0; Mismatches 12; Indels 0; Gaps 0;  
 QY 1 ATTAATTTTACTGCTGGAATACCTTTGTAGAAAACACGAAAGAACTTTCAAGCCTGGG 60  
 DB 520 ACTAATTTTACTGCTGGAATACCTTTGTAGAAAACACGAAAGAACTTTCAAGCCTGGG 579  
 QY 61 AAGGCGCTGATGAATAATTCAGTTGCTCTCTCCAGACAGCTTCGGCGATCCTTTTG 116  
 DB 580 AAGGCGCTGATGAATAATTCAGTTGCTCTCTCCAGACAGCTTCGGCGATCCTTTTG 635  
 XX  
 RESULT 11  
 ABN96785  
 ID ABN96785 standard; DNA; 1348 BP.  
 XX  
 AC ABN96785;  
 XX  
 DT 13-AUG-2002 (first entry)  
 XX



DE Gene #3283 used to diagnose liver cancer.

KW Gene; liver cancer; da; hepatocellular carcinoma; hepatotropic;  
KW metastatic liver tumor; cytostatic; expression profile; disease state;  
XX disease progression; drug toxicity; drug efficacy; drug metabolism.  
XX Homo sapiens.

FN WO200229103-A2.

PD 11-APR-2002.

PP 02-OCT-2001; 2001MO-US030589.

RR 02-OCT-2000; 2000US-0237054P.

XX (GENE-) GENE LOGIC INC.

PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;  
XX WPI; 2002-426119/45.

PT Diagnosing and detecting the progression of liver cancer, hepatocellular  
PT carcinoma or metastatic liver tumor in a patient, involves detecting the  
PT level of expression of two or more genes in a liver tissue sample.

PS Claim 1; SEQ ID NO 3283; 298bp; English.

CC The invention relates to a novel method for diagnosing and detecting the  
CC progression of liver cancer, hepatocellular carcinoma or metastatic liver  
CC tumour in a patient, and differentiating metastatic liver cancer from  
CC hepatocellular carcinoma in a patient, involving detecting the level of  
CC expression of two or more genes represented in ABN93503-ABN97455 in a  
CC tissue sample. The method of the invention has hepatotropic, and  
CC cytosolic activity. The method is useful for diagnosing and detecting  
CC the progression of liver cancer, hepatocellular carcinoma and metastatic  
CC liver carcinoma in a patient. The method is useful for identifying  
CC expression profiles which serve as useful diagnostic markers as well as  
CC markers that can be used to monitor disease states, disease progression,  
CC drug toxicity, drug efficacy and drug metabolism. Note: The sequence data  
CC for this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

SQ Sequence 1348 BP; 370 A; 345 C; 341 G; 291 T; 0 U; 1 Other;

Query March 29.5%; Score 34.2; DB 6; Length 1348;  
Best Local Similarity 58.3%; Pred. No. 0.27; Mismatches 43; Indels 0; Gaps 0  
Matches 60; Conservative 0;

DY 10 ACTGCTGGAACTACTTTGTATAGAAAACCAAGAACAATTTCGAAGCGTGGAGGCTGC 69  
Db 554 ACGTGCTGGGACACCTTTGTGTGACCAACGAGGATGTCCTTCACACCCTGGGATGACTG 613  
DY 70 ATGAAAAATTGAGTTCCTCTCTCCTCAAGACGCTTGGCGCATCTT 112  
Db 614 ATGAGCACACGCCAAGCCTCGACTGTGGAGGCTGGCGGCATCTCT 656

RESULT 12  
ID AAA12409 standard; cDNA, 1446 BP.  
AC AAA12409;  
DT 25-JUL-2000 (first entry)  
DE cDNA encoding a human RNA-associated protein.  
KW Human; RNA-associated protein; cell proliferation; cancer; inflammation;  
KW immune response; reproductive disorder; actinic keratosis;  
KW atherosclerosis; arteriosclerosis; burnsitis; cirrhosis; hepatitis;  
mixed connective tissue disease; myelofibrosis; primary thrombocythemia;

KM		paroxysmal nocturnal hemoglobinuria; polycythemia vera; psoriasis;
KW		trauma; ss.
XX		
OS	Homo sapiens.	
XX		
PH	Key	Location/Qualifiers
PT	CDS	282..1446 /*tag= a /product= "RNA-associated protein"
FT		
XX		
PN	MO20005799-A2.	
PD	23-MAR-2000.	
XX		
PF	17-SEP-1999;	99WO-US021688.
XX		
PR	17-SEP-1998;	98US-00156039.
PR	22-SEP-1998;	98US-00158720.
PR	04-NOV-1998;	98US-00186815.
PR	08-APR-1999;	99US-0128660P.
XX		
PA	(INCY-) INCYTE PHARM INC.	
PI	Tang YT, Corley NC, Guegler KL, Gorgone GA, Patterson C; Hillman JL, Baughn MR, Lai P, Azimzal Y, Yue H, Yang J; MPI; 2000-271437/23. P-P8DB; AAY84437.	
DR		
XX		
PT	New polypeptides and polynucleotides, useful for preventing and treating a disorder associated with increased or decreased expression of RNA associated proteins.	
PS	Claim 9; Page 119; 13pp; English.	
CC	The present sequence encodes a human RNA-associated protein. The expression of RNA-associated proteins is closely associated with reproductive tissues, nervous tissues, cell proliferation including cancer, inflammation and immune responses, and so they may be used for diagnosis, treatment or prevention of cell proliferative, immune/inflammatory disorders, and reproductive disorders. Diseases and disorders which may be treated include actinic keratosis, atherosclerosis, arteriosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease, myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia and cancers, and trauma	
SQ	Sequence 1446 BP; 370 A; 374 C; 393 G; 308 T; 0 U; 1 Other;	
	Query Match 29.5%; Score 34.2; DB 3; Length 1446; Best Local Similarity 58.3%; Pred. No. 0.28; Matches 60; Conservative 0; Mismatches 43; Indels 0; Gaps 0;	
OY	10 ACTGCTGGATACTTTTGTAGAAAAACCAAGAAACAATTTCATAAGCCTGGAGAGGGCTGC 69	
Db	1315 ACTCTGGGACACCTTTGTGGACCACACAGGAGATGCCCTTCCAGCCCTGGATGACTAG 1374	
OY	70 ATGAATAATTAAGTCTGCTCTCCAGACAGCTTGCGGCGCATTCCT 112	
Db	1375 AGGAGCACAGCCCAAGCCCTGATGGTAGGCTGCGGGGCATTTCT 1417	
RESULT 13		
ID	AAZ20856 standard; cDNA; 1534 BP.	
AAZ20856		
AC	AAZ20856;	
XX		
DT	09-DEC-1999 (first entry)	
XX		
DE	Polynucleotide sequence of the Ip547_4 clone.	
XX		
KX	secreted protein; cDNA library; clone; transmembrane protein;	



KM signal sequence cloning; hybridization cloning; gene therapy; receptor;  
 KM ds.  
 XX Homo sapiens.  
 OS  
 FH Key 51.1205 Location/Qualifiers  
 FT CDS  
 FT sig\_peptide /tag= a /product= "Ip547\_4"  
 FT 882..980  
 FT mat\_peptide /tag= b  
 FT 981..1205  
 FT /\*tag= c  
 PN WO9942470-A1.  
 PD 26-AUG-1999.  
 PF 18-FEB-1999; 99WO-US003458.  
 XX  
 PR 18-FEB-1998; 98US-0075038P.  
 PR 17-FEB-1999; 99US-00251600.  
 XX  
 PA (GEMV ) GENETICS INST INC.  
 XX  
 PI Jacobs K, Mccoy JM, Lavallie ER, Collins-Racie LA, Merberg D;  
 PI Treacy M, Agostino MJ, Steininger RJ;  
 DR MPI; 1999-518580/43.  
 DR P-PSDB; AA942383.  
 XX  
 PT New polynucleotides encoding human secreted proteins used for  
 PT therapeutic, diagnostic and research purposes.  
 XX  
 PS Claim 16; Page 104; 125pp; English.  
 XX  
 CC This is the polynucleotide sequence of the clone Ip547\_4, which was  
 CC isolated from a human fetal brain cDNA library using methods which are  
 CC selective for cDNAs encoding secreted proteins, or by identification as a  
 CC secreted or transmembrane protein on the basis of computer analysis of  
 CC the amino acid sequence of the encoded protein. The PNs and proteins of  
 CC the invention are predicted to have biological activities which would  
 CC make them suitable for treating, preventing or ameliorating medical  
 CC conditions in humans and animals, although no supporting data is given.  
 CC Suggested activities include nutritional activity, cytokine and cell  
 CC proliferation/differentiation activity, immune stimulating (e.g. as  
 CC vaccines) or suppressing activity, hematopoiesis regulating activity,  
 CC tissue growth activity, activin/inhibin activity,  
 CC chemotactic/chemokinetic activity, hemostatic and thrombolytic activity,  
 CC receptor/ligand activity, anti-inflammatory activity, cadherin/tumor  
 CC invasion suppressor activity, and tumor inhibition activity. The PNs are  
 CC also stated to be useful for gene therapy. Other activities include  
 CC inhibiting the growth, infection or function of bacteria, fungi, viruses  
 CC and other parasites; effecting bodily characteristics such as, e.g.  
 CC weight, color, skin, etc., effecting biorhythms or circadian cycles;  
 CC enhancing fertility; treatment of depression; treatment of pain; hormonal  
 CC or endocrine activity  
 CC  
 XX  
 SQ Sequence 1534 BP; 495 A; 363 C; 345 G; 331 T; 0 U; 0 Other;  
 XX  
 QY Query Match 29.5%; Score 34.2; DB 2; Length 1534;  
 DB Best Local Similarity 58.3%; Pred. No. 0.29; Indels 0; Gaps 0;  
 DB Matches 60; Conservative 0; Mismatches 43; Indels 0; Gaps 0;  
 QY 10 ACTGCTGGAATCTTTGTAGAAACCAACCAAGCACTTCAAGCTGGAGAGGCTGC 69  
 DB 1084 ACTGCTGGGACACCTTTGTGACACCAAGGAGTGTCCCTCCAGCCCTGGATGACTG 1143  
 QY 70 ATGAATAATTCAGTTCCTCTCCAGACAGCTGGCGCATCTT 112  
 DB 1144 ATGAGCAGACGCAAGACCTGAGTGGGAGGCTGGCGGCATTTCT 1186

RESULT 14  
 AAS59293  
 ID AAS59293 standard; cDNA, 1534 BP.  
 XX  
 AC AAS59293;  
 XX  
 DT 16-JAN-2002 (first entry)  
 XX  
 DE Human cDNA encoding a secreted protein Ip547\_4.  
 XX  
 KW Human; secreted protein; ss; antiinflammatory; immunosuppressive;  
 KW neurotropic; neuroprotective; antidiabetic; antimicrobial; vulnerary;  
 KW cytosolic; antidiabetic; virucide; antiinfectivity; anticonvulsant;  
 KW vasotropic; antiparkinsonian; immunostimulant; dermatological;  
 KW antirheumatic; antitumor; anticancer; osteopathic; tranquilizer;  
 KW cerebroprotective; cytokine; cell proliferation; cell differentiation;  
 KW immune deficiency; severe combined immunodeficiency; SCID; tumor;  
 KW autoimmune disorder; multiple sclerosis; rheumatoid arthritis;  
 KW graft-versus-host disease; myeloid deficiency; wound healing; ulcer;  
 KW periodontal disease; osteoporosis; osteoarthritis; Alzheimer's disease;  
 KW Parkinson's disease; Huntington's disease; infection; cardiac disease;  
 KW stroke; sepsis; inflammatory bowel disease; contraceptive; immunogen;  
 KW food supplement.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175066-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 22-MAR-2001; 2001WO-US009369.  
 XX  
 PR 30-MAR-2000; 2000US-00539330.  
 PR 04-DEC-2000; 2000US-00729674.  
 XX  
 XX  
 PA (GEMV ) GENETICS INST INC.  
 XX  
 PI Jacobs K, Mccoy JM, Lavallie B, Collins-Racie LA, Evans C;  
 PI Treacy M, Agostino MJ, Steininger RJ, Spaulding V, Wong GG, Clark H;  
 PI Pfeitel K, Merberg D;  
 DR MPI; 2001-639363/73.  
 DR P-PSDB; AAU39075.  
 XX  
 PT Secreted human proteins, useful as vaccine for treating various diseases  
 PT such as autoimmune disorders (e.g. multiple sclerosis), and nervous  
 PT system disorders (e.g. stroke).  
 XX  
 PS Disclosure; Page 580; 619pp; English.  
 XX  
 CC The invention relates to novel human secreted proteins, the nucleic acids  
 CC encoding them. The protein may exhibit cytokine, cell proliferation or  
 CC cell differentiation activity or may induce production of other cytokines  
 CC in certain cell populations and may exhibit immune stimulating or immune  
 CC suppressing activity, which is useful for the treatment of various immune  
 CC deficiencies and disorders e.g. severe combined immunodeficiency (SCID),  
 CC autoimmune disorders e.g. multiple sclerosis, systemic lupus  
 CC erythematosus, rheumatoid arthritis, autoimmune pulmonary inflammation.  
 CC The proteins are also useful in the treatment of diseases and disorders  
 CC including tissue, skin and organ transplantation and in graft-versus-host  
 CC diseases (GVHD), in the induction of tumour immunity, myeloid or lymphoid  
 CC cell deficiencies, wound healing and tissue repair, in the treatment of  
 CC burns, incisions and ulcers; as well as in treatment of periodontal  
 CC disease, osteoporosis or osteoarthritis, mediated by inflammatory  
 CC processes, diseases of the peripheral nervous system, Alzheimer's,  
 CC Parkinson's disease, Huntington's disease, amyotrophic lateral  
 CC sclerosis, and Shy-Drager syndrome, infections, infection of cardiac and  
 CC central nervous system vessel e.g. stroke, sepsis, inflammatory bowel  
 CC disease, ulcers, bone regeneration. The protein, having activin- or  
 CC inhibin-related activities is useful as a contraceptive based on the  
 CC ability of inhibin to decrease fertility in female mammals and decrease  
 CC spermatogenesis in male mammals. The proteins and nucleic acids are also  
 CC useful as food supplements. The present sequence encodes a secreted



CC Protein of the invention  
 XX Sequence 1534 BP; 495 A; 363 C; 345 G; 331 T; 0 U; 0 Other;  
 SQ

Query Match 29.5%; Score 34.2; DB 4; Length 1534;  
 Best Local Similarity 58.3%; Pred. No. 0.29;  
 Matches 60; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

10 ACTGCTGGAATCTTTGTATGAAACCCGAAAGACTTTCAAGCCTGGAGGGCTGC 69  
 1084 ACTGCTGGACACCTTTGTGACCAACGAGATGCTCCCTTCAGCCTGGATGAGACTAG 1143

70 ATGAAATTCAGTCTGCTCCAGACAGCTTGGGGCAGTCT 112  
 1144 ATGAGCACAGCCAGACCTGATGAGGAGGCTGCGGGCCATTCT 1186

RESULT 15  
 ABA90962  
 ID ABA90962 standard; cDNA; 1534 BP.  
 AC ABA90962;  
 AT 14-FEB-2002 (first entry)  
 DE Human polynucleotide SEQ ID NO 173.  
 XX Human; clone bd306-7; clone yb8-1; ATCC number 98599; gene therapy;  
 KW immune disorder; bacterial infection; fungal infection; cancer; tumor;  
 KW autoimmune disorder; systemic lupus erythematosus; wound; ulcer; inhibin;  
 KW osteoporosis; osteoarthritis; nervous system disorder; neuropathy;  
 KW Alzheimer's disease; Parkinson's disease; Huntington's disease; activin;  
 KW ischaemia-reperfusion injury; stroke; sepsis; arthritis; vulnery;  
 KW Crohn's disease; cytostatic; anti-inflammatory; immunomodulator;  
 KW neuroprotective; haemostatic; thrombolytic; anti-inflammatory; ss.  
 OS Homo sapiens.  
 XX US2001039335-A1.  
 PN 08-NOV-2001.  
 XX 04-DEC-2000; 2000US-00729674.  
 XX 26-NOV-1997; 97US-0126425P.  
 XX 04-DEC-1997; 97US-0067454P.  
 XX 20-DEC-1997; 97US-0068379P.  
 XX 02-JAN-1998; 98US-0070346P.  
 XX 07-JAN-1998; 98US-0070643P.  
 XX 08-JAN-1998; 98US-0070755P.  
 XX 13-JAN-1998; 98US-0071304P.  
 XX 22-JAN-1998; 98US-0072134P.  
 XX 30-JAN-1998; 98US-0073095P.  
 XX 18-FEB-1998; 98US-0075038P.  
 XX 23-NOV-1998; 98US-00197886.  
 XX 30-MAR-2000; 2000US-00539330.  
 XX (JACO/) JACOBS K.  
 XX (MCCO/) MCCOY J M.  
 XX (LAVA/) LAVALLIE E R.  
 XX (COLL/) COLLINS-RACIE L A.  
 XX (EVAN/) EVANS C.  
 XX (MERB/) MERBERG D.  
 XX (TREA/) TREACY M.  
 XX (AGOS/) AGOSTINO M J.  
 XX (STEL/) STEININGER R J.  
 XX (SPAU/) SPAULDING V.  
 XX (WONG/) WONG G G.  
 XX (CLAR/) CLARK H.  
 XX (FECH/) FECHTEL K.  
 XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;

PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;  
 PI Wong GG, Clark H, Fechtel K;  
 XX WPI: 2002-040725/05.  
 DR P-PSDB: ABB55784.  
 XX

PT New secreted proteins and encoding polynucleotides, useful in gene  
 PT therapies, particularly for preventing or treating autoimmune disorders,  
 PT cancer, graft-versus-host disease, wound, osteoporosis, stroke or  
 PT inflammations.  
 XX

PS Disclosure; Page 310-311; 349pp; English.

XX The invention relates to isolated polynucleotides (ABA90876-ABA90968 and  
 CC ABA90980) and encoded proteins (ABB55698-ABB55800), especially  
 CC polynucleotides SEQ ID NO 1 (ABA90876) and SEQ ID NO 19 (ABA90885) and  
 CC proteins SEQ ID NO 2 (ABB55698) and SEQ ID NO 20 (ABB55707) contained in  
 CC clones bd306-7 and yb8-1 respectively and the clones bd306-7 and yb8-1  
 CC are deposited with the American Type Culture Collection (ATCC) with  
 CC accession number 98599. The polynucleotides and encoded polypeptides have  
 CC cytostatic, anti-inflammatory, immunomodulator, vulnery,  
 CC neuroprotective, activin, inhibin, chemotactic, haemostatic, thrombolytic  
 CC and anti-inflammatory activity and acting as cytokine modulators,  
 CC haematopoiesis regulators, tissue growth modulators and/or cadherin  
 CC suppressors. The polypeptides and polynucleotides are useful in gene  
 CC therapies, particularly for preventing, treating or ameliorating any of  
 CC the following diseases: immune deficiency and disorders; e.g. bacterial  
 CC or fungal infections, autoimmune disorders, cancer, systemic lupus  
 CC erythematosus or graft-versus-host disease, myeloid or lymphoid cell  
 CC deficiencies; wound, burns, incisions and ulcers, osteoporosis or  
 CC osteoarthritis; central and peripheral nervous system diseases and  
 CC neuropathies, e.g. Alzheimer's, Parkinson's disease, Huntington's  
 CC disease, amyotrophic lateral sclerosis or Shy-Drager syndrome;  
 CC haemophilia, cardiac infarction or stroke, inflammations, shock, sepsis  
 CC or systemic inflammatory response syndrome, ischaemia-reperfusion injury,  
 CC endotoxin lethality, arthritis, inflammatory bowel disease or Crohn's  
 CC disease; or tumours or cancers, pemphigus vulgaris or pemphigus foliaceus  
 XX

XX Sequence 1534 BP; 495 A; 363 C; 345 G; 331 T; 0 U; 0 Other;  
 SQ

Query Match 29.5%; Score 34.2; DB 6; Length 1534;  
 Best Local Similarity 58.3%; Pred. No. 0.29;  
 Matches 60; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

10 ACTGCTGGAATCTTTGTATGAAACCCGAAAGACTTTCAAGCCTGGAGGGCTGC 69  
 1084 ACTGCTGGACACCTTTGTGACCAACGAGATGCTCCCTTCAGCCTGGATGAGACTAG 1143

70 ATGAAATTCAGTCTGCTCCAGACAGCTTGGGGCAGTCT 112  
 1144 ATGAGCACAGCCAGACCTGATGAGGAGGCTGCGGGCCATTCT 1186

Search completed: March 12, 2004, 18:08:10  
 Job time : 47.4502 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 12, 2004, 17:39:25 ; Search time 8.63095 Seconds  
(without alignments)  
7458.540 Million cell updates/sec

Title: US-09-966-880A-14

Perfect score: 116  
Sequence: 1 attatcttaccgctggaat.....agcttcgcatcctttg 116

Scoring table: IDENTITY NTC  
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:\*  
1: /cgn2\_6/prodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/prodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/prodata/2/ina/5A\_COMB.seq:\*  
4: /cgn2\_6/prodata/2/ina/5B\_COMB.seq:\*  
5: /cgn2\_6/prodata/2/ina/5A\_COMB.seq:\*  
6: /cgn2\_6/prodata/2/ina/5B\_COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	27.2	23.4	2088	4	US-09-513-057C-4
C 2	27.2	23.4	2518	4	US-09-513-057C-1
C 3	27.2	23.4	2606	4	US-09-513-057C-3
C 4	27.2	23.4	4221	4	US-09-513-057C-3
C 5	27.2	23.4	4221	4	US-09-513-057C-3
C 6	27.2	23.4	8302	4	US-09-513-057C-34
C 7	26.8	23.1	1125	4	US-07-928-462-1
C 8	26.8	23.1	1125	3	US-08-273-247-1
C 9	26.4	22.8	289	4	US-09-513-294A-2945
C 10	26.2	22.6	80246	3	US-09-078-294-4
C 11	26.2	22.6	80595	3	US-09-078-294-3
C 12	26.2	22.6	246	4	US-09-543-681A-3465
C 13	26.2	22.4	4944	1	US-08-623-679-6
C 14	26.2	22.4	4944	1	US-08-623-679-6
C 15	26.2	22.4	4944	3	US-08-623-679-6
C 16	26.2	22.4	4944	3	US-09-181-030-6
C 17	26.2	22.4	4944	4	US-09-534-242-6
C 18	26.2	22.4	4944	4	US-09-534-242-6
C 19	26.2	22.4	4944	4	US-09-164-854-6
C 20	26.2	22.4	5055	3	US-08-623-679-8
C 21	26.2	22.4	5055	3	US-08-623-679-8
C 22	26.2	22.4	5055	3	US-09-181-030-8
C 23	26.2	22.4	5055	4	US-09-534-242-8
C 24	26.2	22.4	5055	4	US-09-534-242-8
C 25	26.2	22.4	5055	4	US-09-164-854-8
C 26	26.2	22.4	5055	4	US-09-164-854-8
C 27	26.2	22.4	5055	4	US-08-637-759B-38
C 28	26.2	22.4	5055	4	US-08-637-759B-38
C 29	26.2	22.4	5055	4	US-08-637-759B-38
C 30	26.2	22.4	5055	4	US-08-637-759B-38
C 31	26.2	22.4	5055	4	US-08-637-759B-38
C 32	26.2	22.4	5055	4	US-08-637-759B-38
C 33	26.2	22.4	5055	4	US-08-637-759B-38
C 34	26.2	22.4	5055	4	US-08-637-759B-38
C 35	26.2	22.4	5055	4	US-08-637-759B-38
C 36	26.2	22.4	5055	4	US-08-637-759B-38
C 37	26.2	22.4	5055	4	US-08-637-759B-38
C 38	26.2	22.4	5055	4	US-08-637-759B-38
C 39	26.2	22.4	5055	4	US-08-637-759B-38
C 40	26.2	22.4	5055	4	US-08-637-759B-38
C 41	26.2	22.4	5055	4	US-08-637-759B-38
C 42	26.2	22.4	5055	4	US-08-637-759B-38
C 43	26.2	22.4	5055	4	US-08-637-759B-38
C 44	26.2	22.4	5055	4	US-08-637-759B-38
C 45	26.2	22.4	5055	4	US-08-637-759B-38

C 28	25.4	21.9	452	1	US-08-158-189-6	Sequence 6, Appli
C 29	25.4	21.9	452	4	US-09-016-434-1198	Sequence 1198, Ap
C 30	25.4	21.9	627	4	US-09-134-001C-1612	Sequence 1612, Ap
C 31	25.4	21.9	1100	3	US-09-248-335-53	Sequence 53, Appl
C 32	25.2	21.7	508	4	US-09-621-17979	Sequence 17979, A
C 33	25.2	21.7	508	4	US-09-453-702B-71	Sequence 71, Appl
C 34	25.2	21.6	1830121	4	US-09-557-884-1	Sequence 1, Appli
C 35	25.2	21.6	1830121	4	US-09-557-884-1	Sequence 1, Appli
C 36	24.8	21.4	339	4	US-09-621-12932	Sequence 12932, A
C 37	24.8	21.4	684	4	US-09-328-352-4049	Sequence 4049, Ap
C 38	24.8	21.4	1001	4	US-09-641-638-131	Sequence 131, App
C 39	24.8	21.4	1318	4	US-09-443-041A-25	Sequence 25, Appl
C 40	24.8	21.4	5763	4	US-08-956-171E-505	Sequence 505, App
C 41	24.6	21.2	416	4	US-09-621-18848	Sequence 18848, A
C 42	24.6	21.2	567	4	US-09-621-976-1240	Sequence 1240, Ap
C 43	24.6	21.2	1664976	4	US-08-916-421B-1	Sequence 1, Appli
C 44	24.4	21.0	207	4	US-09-134-001C-21	Sequence 21, Appl
C 45	24.4	21.0	511	4	US-09-621-976-3532	Sequence 3532, Ap

## ALIGNMENTS

RESULT 1  
US-09-513-057C-4/c  
Sequence 4, Application US/09513057C  
Patent No. 6433251

GENERAL INFORMATION:

APPLICANT: Wagner, et al.

TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM

FILE REFERENCE: 1505-54357

CURRENT FILING DATE: 2000-02-24

CURRENT FILING DATE: 2000-02-24

NUMBER OF SEQ ID NOS: 35

SOFTWARE: Patentin version 3.1

SEQ ID NO 4

LENGTH: 2088

TYPE: DNA

ORGANISM: Arabidopsis thaliana

US-09-513-057C-4

Query Match 23.4%; Score 27.2; DB 4; Length 2088;  
Best Local Similarity 52.7%; Pred. NO. 3.5;  
Matches 59; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 3 TATTTTACTGCTGGAATACCTTTGTGAAACCAAGAACTTCAAGCCTGGAA 62

DB 184 TGTGTACTACGAGATTCATCTTCATGATGATCAACCTCTGGAAGAACTACTA 125

QY 63 GGGCTGATGAAATTCAGTTCGTCGCAAGAGCTTGGCGCATCTTT 114

DB 124 GCTGCTATAAAGAGCATCTGTTTCTAGAGAGAGCTTAGGCGCTCTTT 73

RESULT 2  
US-09-513-057C-1/c  
Sequence 1, Application US/09513057C  
Patent No. 6433251  
GENERAL INFORMATION:  
APPLICANT: Wagner, et al.  
TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM  
FILE REFERENCE: 1505-54357  
CURRENT FILING DATE: 2000-02-24  
CURRENT FILING DATE: 2000-02-24  
NUMBER OF SEQ ID NOS: 35  
SOFTWARE: Patentin version 3.1  
SEQ ID NO 1  
LENGTH: 2518  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
FEATURES:  
NAME/KEY: CDS  
LOCATION: (284)..(2371)



OTHER INFORMATION:  
US-09-513-057C-1

Query Match 23.4%; Score 27.2; DB 4; Length 2518;  
Best Local Similarity 52.7%; Pred. No. 3.8;  
Matches 59; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 3 TATTTTACTGCTGGAATCTTTGTGAGAAAACCAAGAAAGAACTTCAAGCCTGGGA 62  
DB 467 TGTGTACTACGAGAAATTCATGTCATGATCACCAACCTCTGAGAAGAAATACTTA 408  
QY 63 GGGCTGATGAATAATTCAGTTCTCTCCAGACAGCTTCGGGCACTCTTT 114  
DB 407 GCTGCTCATTAAGAGCCATCTTGTTCCTAGAGAGAGCTCTAGGCCCTCTTT 356

## RESULT 3

US-09-234-827B-3/C  
Sequence 3, Application US/09234827B

PATENT No. 6448471  
GENERAL INFORMATION:  
APPLICANT: Puzio, Piotr S.  
TITLE OF INVENTION: Nematode feeding structure specific gene and its  
FILE REFERENCE: U-012084-2  
CURRENT APPLICATION NUMBER: US/09/234,827B  
CURRENT FILING DATE: 1999-01-21  
PRIORITY APPLICATION NUMBER: US 60/072,142  
PRIORITY FILING DATE: 1998-01-22  
NUMBER OF SEQ ID NOS: 32  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 3  
LENGTH: 2606  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (328)..(2412)  
US-09-234-827B-3

Query Match 23.4%; Score 27.2; DB 4; Length 2606;  
Best Local Similarity 52.7%; Pred. No. 3.9;  
Matches 59; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 3 TATTTTACTGCTGGAATCTTTGTGAGAAAACCAAGAAAGAACTTCAAGCCTGGGA 62  
DB 511 TGTGTACTACGAGAAATTCATGTCATGATCACCAACCTCTGAGAAGAAATACTTA 452  
QY 63 GGGCTGATGAATAATTCAGTTCTCTCCAGACAGCTTCGGGCACTCTTT 114  
DB 451 GCTGCTCATTAAGAGCCATCTTGTTCCTAGAGAGAGCTCTAGGCCCTCTTT 400

## RESULT 4

US-09-513-057C-3/C  
Sequence 3, Application US/09513057C

PATENT No. 6433251  
GENERAL INFORMATION:  
APPLICANT: Wagner, et al.  
TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM  
FILE REFERENCE: 1505-54357  
CURRENT APPLICATION NUMBER: US/09/513,057C  
CURRENT FILING DATE: 2000-02-24  
NUMBER OF SEQ ID NOS: 35  
SOFTWARE: Patentin version 3.1  
SEQ ID NO 3  
LENGTH: 4221  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
FEATURE:  
NAME/KEY: promoter  
LOCATION: (1)..(142)

OTHER INFORMATION:  
NAME/KEY: 5'UTR  
LOCATION: (143)..(425)  
OTHER INFORMATION:  
NAME/KEY: exon  
LOCATION: (426)..(644)  
OTHER INFORMATION:  
NAME/KEY: exon  
LOCATION: (1007)..(1803)  
OTHER INFORMATION:  
NAME/KEY: exon  
LOCATION: (2984)..(3037)  
OTHER INFORMATION:  
NAME/KEY: exon  
LOCATION: (3128)..(4142)  
OTHER INFORMATION:  
NAME/KEY: Intron  
LOCATION: (645)..(1006)  
OTHER INFORMATION:  
NAME/KEY: Intron  
LOCATION: (1804)..(2983)  
OTHER INFORMATION:  
NAME/KEY: Intron  
LOCATION: (3038)..(3127)  
OTHER INFORMATION:  
NAME/KEY: 3'UTR  
LOCATION: (4146)..(4221)  
OTHER INFORMATION:  
US-09-513-057C-3

Query Match 23.4%; Score 27.2; DB 4; Length 4221;  
Best Local Similarity 52.7%; Pred. No. 4.8;  
Matches 59; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 3 TATTTTACTGCTGGAATCTTTGTGAGAAAACCAAGAAAGAACTTCAAGCCTGGGA 62  
DB 609 TGTGTACTACGAGAAATTCATGTCATGATCACCAACCTCTGAGAAGAAATACTTA 550  
QY 63 GGGCTGATGAATAATTCAGTTCTCTCCAGACAGCTTCGGGCACTCTTT 114  
DB 549 GCTGCTCATTAAGAGCCATCTTGTTCCTAGAGAGAGCTCTAGGCCCTCTTT 498

## RESULT 5

US-09-513-057C-34/C  
Sequence 34, Application US/09513057C

PATENT No. 6433251  
GENERAL INFORMATION:  
APPLICANT: Wagner, et al.  
TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM  
FILE REFERENCE: 1505-54357  
CURRENT APPLICATION NUMBER: US/09/513,057C  
CURRENT FILING DATE: 2000-02-24  
NUMBER OF SEQ ID NOS: 35  
SOFTWARE: Patentin version 3.1  
SEQ ID NO 34  
LENGTH: 4221  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
FEATURE:  
NAME/KEY: 5'UTR  
LOCATION: (1)..(425)  
OTHER INFORMATION:  
NAME/KEY: exon  
LOCATION: (426)..(644)  
OTHER INFORMATION:  
NAME/KEY: Intron  
LOCATION: (645)..(1006)  
OTHER INFORMATION:  
NAME/KEY: exon  
LOCATION: (1007)..(1803)  
OTHER INFORMATION:  
NAME/KEY: Intron



LOCATION: (1804)..(2983)  
OTHER INFORMATION:  
NAME/KEY: exon  
LOCATION: (2984)..(3035)  
OTHER INFORMATION:  
NAME/KEY: Intron  
LOCATION: (3036)..(3125)  
OTHER INFORMATION:  
NAME/KEY: exon  
LOCATION: (3126)..(4145)  
OTHER INFORMATION:  
NAME/KEY: 3'UTR  
LOCATION: (4146)..(4221)  
OTHER INFORMATION:  
US-09-513-057C-34

Query Match 23.4%; Score 27.2; DB 4; Length 4221;  
Best Local Similarity 52.7%; Pred. No. 4.8;  
Matches 59; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 3 TATTTTACTGCTGGAATCTTTGTAGAAACACGAAAGAACTTCAAGCCTGGAA 62  
DB 609 TGTGTACTACGAGAAATCTCTTCATGATCACCACCTTGAAGAGAACTCTAA 550  
QY 63 GGGCTGATGAAATTCAGTTCCTCTCCAGACGCTGGCGCATCTTT 114  
DB 549 GCTGCTATTAAGAGCATCTTGTCTTAGAGAGAGCTTAGGCCCTCTTT 498

RESULT 6  
US-09-234-827B-1/c  
Sequence 1, Application US/09234827B  
Patent No. 6448471  
GENERAL INFORMATION:  
APPLICANT: Puzio, Piotr S.  
APPLICANT: Grunfelder, Florian M.W.  
TITLE OF INVENTION: Nematode feeding structure specific gene and its  
FILE REFERENCE: U-012064-2  
CURRENT APPLICATION NUMBER: US/09/234,827B  
CURRENT FILING DATE: 1999-01-21  
PRIOR APPLICATION NUMBER: US 60/072,142  
PRIOR FILING DATE: 1998-01-22  
NUMBER OF SEQ ID NOS: 32  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 1  
LENGTH: 8302  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (4038)..(4256)  
NAME/KEY: CDS  
LOCATION: (4807)..(5604)  
NAME/KEY: CDS  
LOCATION: (6777)..(6827)  
NAME/KEY: CDS  
LOCATION: (6919)..(7935)  
US-09-234-827B-1

Query Match 23.4%; Score 27.2; DB 4; Length 8302;  
Best Local Similarity 52.7%; Pred. No. 6.5;  
Matches 59; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 3 TATTTTACTGCTGGAATCTTTGTAGAAACACGAAAGAACTTCAAGCCTGGAA 62  
DB 4221 TGTGTACTACGAGAAATCTCTTCATGATCACCACCTTGAAGAGAACTCTAA 4162  
QY 63 GGGCTGATGAAATTCAGTTCCTCTCCAGACGCTGGCGCATCTTT 114  
DB 4161 GCTGCTATTAAGAGCATCTTGTCTTAGAGAGAGCTTAGGCCCTCTTT 4110

## RESULT 7

US-07-928-462-1  
Sequence 1, Application US/07928462  
Patent No. 5328996

## GENERAL INFORMATION:

APPLICANT: Boyle, Michael D.P.  
APPLICANT: Lottenberg, Richard  
APPLICANT: Broder, Christopher C.  
APPLICANT: von Merling, Gregory O.  
TITLE OF INVENTION: Bacterial plasmin Receptors as  
TITLE OF INVENTION: Fibrinolytic Agents  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: David R. Saliwanichik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/928,462  
FILING DATE: 19920810

## CLASSIFICATION: 424

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/524,411  
FILING DATE: 16-MAY-1990  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/330,849  
FILING DATE: 29-MAR-1989

## ATTORNEY/AGENT INFORMATION:

NAME: Saliwanichik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: UF/S&S-13.C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800

## INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:  
LENGTH: 1125 base pairs  
TYPE: NUCLEIC ACID

STRANDEDNESS: both  
TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO

FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:

ORGANISM: Streptococcus pyogenes  
STRAIN: M untypable

INDIVIDUAL ISOLATE: 64/14  
IMMEDIATE SOURCE:

CLONE: pRL015  
FEATURE:

NAME/KEY: CDS  
LOCATION: 115..1122

FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 115..1122

IDENTIFICATION METHOD: experimental  
OTHER INFORMATION: /codon start= 115

OTHER INFORMATION: /function= "High-affinity binding of  
plasmin(ogen)"

OTHER INFORMATION: /product= "Streptococcal plasmin receptor"  
OTHER INFORMATION: /evidence= EXPERIMENTAL

OTHER INFORMATION: /gene= "plr"  
OTHER INFORMATION: /number= 1

OTHER INFORMATION: /label= plr  
US-07-928-462-1



Query Match 23.1%; Score 26.8; DB 1; Length 1125;  
Best Local Similarity 59.0%; Pred. No. 3.7;  
Matches 46; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 35 CCGAAGAAAGCTTCAAGCCCTGGAGAGCGCTGCATGAATAATTCAGTTCCTCCAG 94  
DB 899 CTATGAAGCTGCTTCAACGATAGCTTCGTTACCTGAAGATCCATCGTTCTTCAG 958

QY 95 ACAGCTCGGCGCATCCT 112  
DB 959 ATATCGTAGCGGTATCAT 976

US-08-273-247-1  
Sequence 1, Application US/08273247  
Patent No. 6136323

GENERAL INFORMATION:  
APPLICANT: Boyle, Michael D.P.  
APPLICANT: Lotenberg, Richard  
APPLICANT: Broder, Christopher C.  
APPLICANT: von Mering, Gregory O.  
TITLE OF INVENTION: Bacterial Plasmin Receptors as  
NUMBER OF INVENTION: Fibrinolytic Agents  
TITLE OF INVENTION: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David R. Saliwanchik  
STREET: 2421 N.W. 41st Street, Suite A-1  
CITY: Gainesville  
STATE: FL  
COUNTRY: USA  
ZIP: 32606

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/273,247  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/524,411  
FILING DATE: 16-MAY-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/330,849  
FILING DATE: 29-MAR-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Saliwanchik, David R.  
REGISTRATION NUMBER: 31,794  
REFERENCE/DOCKET NUMBER: UF/S&S-13.C2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1125 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Streptococcus pyogenes  
STRAIN: M untypable  
INDIVIDUAL ISOLATE: 64/14  
IMMEDIATE SOURCE:  
CLONE: pRL015  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 115..1122  
FEATURE:

NAME/KEY: mac peptide  
LOCATION: 115..1122  
IDENTIFICATION METHOD: experimental  
OTHER INFORMATION: /codon\_start=115  
OTHER INFORMATION: /function="High-affinity binding of  
OTHER INFORMATION: /plasma(ogen)"  
OTHER INFORMATION: /product="Streptococcal plasmin receptor"  
OTHER INFORMATION: /evidence=EXPERIMENTAL  
OTHER INFORMATION: /gene="plr"  
OTHER INFORMATION: /number=1  
OTHER INFORMATION: /label=PLR

US-08-273-247-1  
Query Match 23.1%; Score 26.8; DB 3; Length 1125;  
Best Local Similarity 59.0%; Pred. No. 3.7;  
Matches 46; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 35 CCGAAGAAAGCTTCAAGCCCTGGAGAGCGCTGCATGAATAATTCAGTTCCTCCAG 94  
DB 899 CTATGAAGCTGCTTCAACGATAGCTTCGTTACCTGAAGATCCATCGTTCTTCAG 958

QY 95 ACAGCTCGGCGCATCCT 112  
DB 959 ATATCGTAGCGGTATCAT 976

US-09-313-294A-2945  
Sequence 2945, Application US/09313294A  
Patent No. 6476212

GENERAL INFORMATION:  
APPLICANT: Lalgudi, Raghunath V.  
APPLICANT: Ito, Laura Y.  
APPLICANT: Sherman, Bradley K.  
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR  
FILE REFERENCE: PL-0017 US  
CURRENT APPLICATION NUMBER: US/09/313,294A  
CURRENT FILING DATE: 1999-05-14  
NUMBER OF SEQ ID NOS: 7600  
SOFTWARE: PERL Program  
SEQ ID NO 2945  
LENGTH: 289  
TYPE: DNA  
ORGANISM: Zea mays  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: incyte ID No. 6476212 700282045H1  
NAME/KEY: unsure  
LOCATION: 22, 89  
OTHER INFORMATION: a, c, g, or other  
US-09-313-294A-2945

Query Match 22.8%; Score 26.4; DB 4; Length 289;  
Best Local Similarity 52.9%; Pred. No. 2.9;  
Matches 54; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 7 TTACTGCTGAATACCTTTGTGAGAAAACACGAGAAATTTCAAGCCTGGAGAGGC 66  
DB 6 TATATCCATGAATACGCGGTAGAGAAAGAGAGATGATGCTCGAAGCAGGGT 65

QY 67 TGCAAGAAATTCAGTTCCTCCGAGAGAGCTGGCGCC 108  
DB 66 AGAGACATACATAAGGCTTCACACGAGAGCTTTGCTCA 107

US-09-078-294-4  
Sequence 4, Application US/09078294  
Patent No. 6265211

GENERAL INFORMATION:  
APPLICANT: Choo, Kong-Hong Andy  
APPLICANT: Du Sart, Desiree  
APPLICANT: Cancilla, Michael R.



;; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE  
;; FILE REFERENCE: Davies Col  
;; CURRENT APPLICATION NUMBER: US/09/078,294  
;; CURRENT FILING DATE: 1998-05-13  
;; NUMBER OF SEQ ID NOS: 29  
;; SOFTWARE: Patentin Ver. 2.0  
;; SEQ ID NO 4  
;; LENGTH: 80246  
;; TYPE: DNA  
;; ORGANISM: Nucleotide sequence of NC-contig  
US-09-078-294-4  
  
Query Match 22.6%; Score 26.2; DB 3; Length 80246;  
Best Local Similarity 52.3%; Pred. No. 41;  
Matches 58; Conservative 0; Mismatches 53; Indels 0; Gaps 0;  
  
QY 5 TTTTCTGCTGGAATCTTTGTAGAAAACGAAAGACTTTGAAGCCTGGGAGG 64  
DB 67330 TTCTTTCTCCAGTAGCATATGTCATATCTTTAAAGAGAAAGAGGGGAGC 67389  
QY 65 GCTGCATGAATAATTCAGTTGCTCTCCAGACAGCTTCGGCGCATCTTTT 115  
DB 67390 ACTGAACCAATTGCTTTTGTACTGCTCAGCTCAATGAGAGTTCT 67440  
  
RESULT 11  
US-09-078-294-3  
; Sequence 3, Application US/09078294  
; Patent No. 6265211  
; GENERAL INFORMATION:  
; APPLICANT: Choo, Kong-Hong Andy  
; APPLICANT: Du Sart, Desiree  
; APPLICANT: Cancilla, Michael R.  
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE  
; FILE REFERENCE: Davies Col  
; CURRENT APPLICATION NUMBER: US/09/078,294  
; CURRENT FILING DATE: 1998-05-13  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 80595  
; TYPE: DNA  
; ORGANISM: Nucleotide sequence of HC-contig  
US-09-078-294-3  
  
Query Match 22.6%; Score 26.2; DB 3; Length 80595;  
Best Local Similarity 52.3%; Pred. No. 41;  
Matches 58; Conservative 0; Mismatches 53; Indels 0; Gaps 0;  
  
QY 5 TTTTCTGCTGGAATCTTTGTAGAAAACGAAAGACTTTGAAGCCTGGGAGG 64  
DB 67588 TTCTTTCTCCAGTAGCATATGTCATATCTTTAAAGAGAAAGAGGGGAGC 67647  
QY 65 GCTGCATGAATAATTCAGTTGCTCTCCAGACAGCTTCGGCGCATCTTTT 115  
DB 67648 ACTGAACCAATTGCTTTTGTACTGCTCAGCTCAATGAGAGTTCT 67698  
  
RESULT 12  
US-09-543-681A-3465/c  
; Sequence 3465, Application US/09543681A  
; Patent No. 6605709  
; GENERAL INFORMATION:  
; APPLICANT: GARY BRETTON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709, 1002-001  
; CURRENT APPLICATION NUMBER: US/09/543,681A  
; CURRENT FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/128,706  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 3465

;; LENGTH: 246  
;; TYPE: DNA  
;; ORGANISM: Proteus mirabilis  
US-09-543-681A-3465  
  
Query Match 22.4%; Score 26; DB 4; Length 246;  
Best Local Similarity 62.1%; Pred. No. 3.8;  
Matches 41; Conservative 0; Mismatches 25; Indels 0; Gaps 0;  
  
QY 31 AAAACGAGAAAGACTTTCAGAGCTGGGAGGCTGCATGAATAATTCAGTTCT 90  
DB 239 AAGAAATCAATAATTAATTAATGACCGATTAAATGTGAATTAATACTGATCTTCA 180  
QY 91 CCAGAC 96  
DB 179 CTGGAC 174  
  
RESULT 13  
US-08-623-679-6  
; Sequence 6, Application US/08623679  
; Patent No. 5674739  
; GENERAL INFORMATION:  
; APPLICANT: Shyjan, Andrew W.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
; TITLE OF INVENTION: DIAGNOSIS, PREVENTION AND TREATMENT OF TUMOR  
; TITLE OF INVENTION: PROGRESSION  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/623,679  
; FILING DATE: 29-MAR-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/412,431  
; FILING DATE: 29-MAR-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fasse, J. Peter  
; REGISTRATION NUMBER: 32,983  
; REFERENCE/DOCKET NUMBER: 07334/004001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4944 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 346..4837  
US-08-623-679-6  
  
Query Match 22.4%; Score 26; DB 1; Length 4944;  
Best Local Similarity 54.1%; Pred. No. 14;  
Matches 53; Conservative 0; Mismatches 45; Indels 0; Gaps 0;  
  
QY 1 ATATTTTCTGCTGGAATCTTTGTAGAAAACGAAAGACTTTGAAGCCTGGG 60



Db 2441 ATGATGATTCTCGTATCAACATCCAGGAAACGAGATGCAAGAAAAGAAAGAGG 2500  
 QY 61 AAGGCTGCATGAAATTCAGTTCGTCTCTCCGACAG 98  
 Db 2501 AAATACGATGCAATGCAGATGCTGCTCAAGAAAG 2538

RESULT 14  
 US-08-933-774-6  
 ; Sequence 6, Application US/08933774A  
 ; Patent No. 6025137  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Shyjan, Andrew W.  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS, PREVENTION  
 ; FILE REFERENCE: 07334/004003  
 ; CURRENT APPLICATION NUMBER: US/08/933,774A  
 ; EARLIER FILING DATE: 1997-09-19  
 ; EARLIER APPLICATION NUMBER: US 08/623,679  
 ; EARLIER FILING DATE: 1996-03-29  
 ; EARLIER APPLICATION NUMBER: US 08/412,431  
 ; NUMBER OF SEQ ID NOS: 10  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 6  
 ; LENGTH: 4944  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (346)...(4836)  
 US-08-933-774-6

Query Match 22.4%; Score 26; DB 3; Length 4944;  
 Best Local Similarity 54.1%; Pred. No. 14;  
 Matches 53; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 1 ATTATTTTACTGCTGGAATCTTTGTAGAAAACCAAGAACTTTCAAGCCTGGG 60  
 Db 2441 ATGATGATTCTCGTATCAACATCCAGGAAACGAGATGCTCAAGAAAAGAGG 2500  
 QY 61 AAGGCTGCATGAAATTCAGTTCGTCTCTCCGACAG 98  
 Db 2501 AAATACGATGCAATGCAGATGCTGCTCAAGAAAG 2538

RESULT 15  
 US-09-181-030-6  
 ; Sequence 6, Application US/09181030  
 ; Patent No. 6251597  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Shyjan, Andrew W.  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS, PREVENTION  
 ; FILE REFERENCE: 07334/004005  
 ; CURRENT APPLICATION NUMBER: US/09/181,030  
 ; EARLIER FILING DATE: 1998-10-27  
 ; EARLIER APPLICATION NUMBER: US 08/862,442  
 ; EARLIER FILING DATE: 1997-05-23  
 ; EARLIER APPLICATION NUMBER: US 08/623,679  
 ; EARLIER FILING DATE: 1996-03-29  
 ; NUMBER OF SEQ ID NOS: 10  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 6  
 ; LENGTH: 4944  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (346)...(4836)  
 US-09-181-030-6

Query Match 22.4%; Score 26; DB 3; Length 4944;

Best Local Similarity 54.1%; Pred. No. 14;  
 Matches 53; Conservative 0; Mismatches 45; Indels 0; Gaps 0;  
 QY 1 ATTATTTTACTGCTGGAATCTTTGTAGAAAACCAAGAACTTTCAAGCCTGGG 60  
 Db 2441 ATGATGATTCTCGTATCAACATCCAGGAAACGAGATGCTCAAGAAAAGAGG 2500  
 QY 61 AAGGCTGCATGAAATTCAGTTCGTCTCTCCGACAG 98  
 Db 2501 AAATACGATGCAATGCAGATGCTGCTCAAGAAAG 2538

Search completed: March 13, 2004, 00:51:23  
 Job time : 11.631 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 12, 2004, 21:22:51 ; Search time 40.7067 Seconds

(without alignments)  
10491.276 Million cell updates/sec

Title: US-09-966-880a-14

Perfect score: 116  
Sequence: 1 attatttactgctgcaatc.....agcttcggcgcacccctttg 116

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2432557 seqs, 1840798884 residues

Total number of hits satisfying chosen parameters: 4865114

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA.\*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US09D\_PUBCOMB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/2/pubpna/US10D\_PUBCOMB.seq.\*  
17: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
18: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	116	100.0	116	US-09-966-880A-14	Sequence 14, Appl
2	116	100.0	6564	US-09-966-880A-10	Sequence 10, Appl
3	116	100.0	11204	US-09-966-880A-35	Sequence 35, Appl
4	114.4	96.6	2818	US-09-966-880A-7	Sequence 7, Appl
5	96.8	83.4	2440	US-09-966-880A-1	Sequence 1, Appl
6	34.2	29.5	390	US-09-796-692-9051	Sequence 9051, Ap
7	34.2	29.5	390	US-10-040-862-9051	Sequence 9051, Ap
8	34.2	29.5	390	US-10-057-475B-9051	Sequence 9051, Ap
9	34.2	29.5	390	US-10-154-884B-9051	Sequence 9051, Ap
10	34.2	29.5	1155	US-10-460-923-1	Sequence 1, Appl
11	34.2	29.5	1348	US-09-880-107-3282	Sequence 3282, Ap
12	34.2	29.5	1534	US-09-729-674-173	Sequence 173, App
13	34.2	29.5	20143	US-10-240-425-1099	Sequence 1099, Ap
14	30.6	26.4	274	US-10-424-559-23145	Sequence 23145, A
15	30.4	26.2	328	US-09-796-692-5737	Sequence 5737, Ap

16	30.4	26.2	328	14	US-10-040-862-5737	Sequence 5737, Ap
17	30.4	26.2	328	15	US-10-057-475B-5737	Sequence 5737, Ap
18	30.4	26.2	328	15	US-10-154-884B-5737	Sequence 5737, Ap
19	30.4	26.2	606	15	US-10-027-632-273902	Sequence 273902, Ap
20	30.4	26.2	823	14	US-10-247-671-62	Sequence 62, Appl
21	30.4	26.2	1008	15	US-10-131-827-8890	Sequence 8890, Ap
22	30.4	26.2	1073	12	US-10-425-114-30021	Sequence 30021, A
23	30.4	26.2	1161	12	US-10-424-599-19811	Sequence 19811, A
24	30.4	26.2	2913	15	US-10-027-632-112023	Sequence 112023, A
25	30.4	26.2	2933	15	US-10-027-632-112024	Sequence 112024, A
26	30.2	26.0	945	15	US-10-398-221-2508	Sequence 2508, Ap
27	30.2	26.0	945	15	US-10-398-221-2571	Sequence 2571, Ap
28	30.2	26.0	966	15	US-10-398-221-2571	Sequence 553, App
29	30.2	26.0	966	15	US-10-398-221-2554	Sequence 554, App
30	30.2	26.0	684707	15	US-10-398-221-9	Sequence 9, Appl
31	30.2	26.0	684707	15	US-10-398-221-9	Sequence 2058, Ap
32	30.2	26.0	3011208	15	US-10-027-632-231323	Sequence 231323, A
33	29.2	25.2	637	15	US-10-027-632-231323	Sequence 231324, A
34	29.2	25.2	637	15	US-10-027-632-231324	Sequence 231325, A
35	29.2	25.2	637	15	US-10-027-632-231325	Sequence 8363, Ap
36	29.2	25.2	767	15	US-10-027-632-8363	Sequence 8363, Ap
37	29.2	25.2	767	15	US-10-027-632-8364	Sequence 33857, A
38	29.2	25.2	3372	12	US-10-425-114-33257	Sequence 205700, A
39	28.8	24.8	663	15	US-10-027-632-64046	Sequence 64046, A
40	28.8	24.7	514	15	US-10-027-632-64046	Sequence 64047, A
41	28.6	24.7	514	15	US-10-027-632-64047	Sequence 45937, A
42	28.6	24.7	668	15	US-10-027-632-45937	Sequence 31, Appl
43	28.6	24.7	68223	15	US-10-034-650-31	Sequence 38035, A
44	28.4	24.5	485	12	US-10-085-783A-38035	Sequence 38035, A
45	28.4	24.5	485	15	US-10-242-535A-38035	Sequence 38035, A

#### ALIGNMENTS

RESULT 1					
US-09-966-880A-14					
Sequence 14, Application US/09966880A					
Patent No. US20020164743A1					
GENERAL INFORMATION:					
APPLICANT: Honjo, Tasuku					
APPLICANT: Muramatsu, Masamichi					
TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE					
FILE REFERENCE: 06501-086001 US/09/966, 880A					
CURRENT APPLICATION NUMBER: US/09/966, 880A					
CURRENT FILING DATE: 2001-09-28					
PRIOR APPLICATION NUMBER: PCT/JP00/01918					
PRIOR FILING DATE: 2000-03-28					
PRIOR APPLICATION NUMBER: JP 11-371382					
PRIOR FILING DATE: 1999-12-27					
PRIOR APPLICATION NUMBER: JP 11-178999					
PRIOR FILING DATE: 1999-06-24					
PRIOR APPLICATION NUMBER: JP 11-87192					
PRIOR FILING DATE: 1999-03-29					
NUMBER OF SEQ ID NOS: 36					
SOFTWARE: FastSeq for Windows Version 4.0					
SEQ ID NO 14					
LENGTH: 116					
TYPE: DNA					
ORGANISM: Homo sapiens					
US-09-966-880A-14					
Query Match					
Best Local Similarity 100.0%; Score 116; DB 9; Length 116;					
Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	1	ATTATTTTACTGCTGGAATCTTTGTGAAACACCAAGAACTTTCAAGCTGGG	60		
DB	1	ATTATTTTACTGCTGGAATCTTTGTGAAACACCAAGAACTTTCAAGCTGGG	60		
QY	61	AAGGGCTGATGAAATTCAGTTCGTCTCCAGACAGCTTGGGGCAGCTTTTG	116		
DB	61	AAGGGCTGATGAAATTCAGTTCGTCTCCAGACAGCTTGGGGCAGCTTTTG	116		



```
RESULT 2
US-09-966-880A-10
; Sequence 10, Application US/09966880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; APPLICANT: Muramatsu, Masamichi
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966,880A
; PRIOR FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 6564
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-966-880A-10

Query Match      100.0%; Score 116; DB 9; Length 6564;
Best Local Similarity 100.0%; Pred. No. 1.1e-30;
Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ATTATTTTACTGCTGGAATTAATTTGTAGAAACACGAAAGAACTTCAAGCTGGG 60
DB      3156 ATTATTTTACTGCTGGAATTAATTTGTAGAAACACGAAAGAACTTCAAGCTGGG 3215

QY      61 AAGGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTGGGCAATCCCTTTTG 116
DB      3216 AAGGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTGGGCAATCCCTTTTG 3271

RESULT 3
US-09-966-880A-35
; Sequence 35, Application US/09966880A
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; APPLICANT: Muramatsu, Masamichi
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966,880A
; PRIOR FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 35
; LENGTH: 11204
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-966-880A-35

Query Match      100.0%; Score 116; DB 9; Length 11204;
Best Local Similarity 100.0%; Pred. No. 1.4e-30;
Matches 116; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 ATTATTTTACTGCTGGAATTAATTTGTAGAAACACGAAAGAACTTCAAGCTGGG 60
DB      8371 ATTATTTTACTGCTGGAATTAATTTGTAGAAACACGAAAGAACTTCAAGCTGGG 8430

QY      61 AAGGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTGGGCAATCCCTTTTG 116
DB      8431 AAGGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTGGGCAATCCCTTTTG 8486

RESULT 4
US-09-966-880A-7
; Sequence 7, Application US/09966880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; APPLICANT: Muramatsu, Masamichi
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966,880A
; PRIOR FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 2818
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (80)...(673)
; NAME/KEY: 5' UTR
; LOCATION: (1)...(79)
; FEATURE:
; NAME/KEY: 3' UTR
; LOCATION: (677)...(2818)
US-09-966-880A-7

Query Match      98.6%; Score 114.4; DB 9; Length 2818;
Best Local Similarity 99.1%; Pred. No. 3e-30;
Matches 115; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 ATTATTTTACTGCTGGAATTAATTTGTAGAAACACGAAAGAACTTCAAGCTGGG 60
DB      507 ATTATTTTACTGCTGGAATTAATTTGTAGAAACACGAAAGAACTTCAAGCTGGG 566

QY      61 AAGGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTGGGCAATCCCTTTTG 116
DB      567 AAGGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTGGGCAATCCCTTTTG 622

RESULT 5
US-09-966-880A-1
; Sequence 1, Application US/09966880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; APPLICANT: Muramatsu, Masamichi
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966,880A
; PRIOR FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
```



```

; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2440
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (93)...(686)
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1)...(92)
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: (690)...(2440)
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(2440)
; OTHER INFORMATION: n = A,T,C or G
US-09-966-880A-1

Query Match      83.4%; Score 96.8; DB 9; Length 2440;
Best Local Similarity 89.7%; Pred. No. 5.5e-24;
Matches 104; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY      1 ATTATTTTACTGCTGGAATCTTTTGTAAGAAACCAAGAACTTCAAGGCTGGG 60
DB      520 ACTATTTTACTGCTGGAATCTTTTGTAAGAAACCAAGAACTTCAAGGCTGGG 579
        |||
QY      61 AAGGCTGATGAATTCAGTTCTCTCCAGAGAGTTGGGCGATCCTTTTG 116
DB      580 AAGGCTGATGAATTCAGTTCTCTCCAGAGAGTTGGGCGATCCTTTTG 635
        |||

RESULT 6
US-09-796-692-9051
; Sequence 9051, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; PRIOR FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
```

```

; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9051
; LENGTH: 390
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-796-692-9051

Query Match      29.5%; Score 34.2; DB 9; Length 390;
Best Local Similarity 58.3%; Pred. No. 0.062;
Matches 60; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY      10 ACTGCTGAATCTTTTGTAAGAAACCAAGAACTTCAAGGCTGGGAGGCTGC 69
DB      165 ACTGCTGAACCTTTGTGACACACAGGATGTCCCTCCAGCCCTGGATGACTAG 224
        |||
QY      70 ATGAAATTCAATTCAGTTCTCTCCAGACAGCTTGGCGCATCCT 112
DB      225 ATGAGCACAGCCCAAGACCTGATGAGGAGCTGCGGCGCATTCCT 267
        |||

RESULT 7
US-10-040-862-9051
; Sequence 9051, Application US/10040862
; Publication No. US20030078396A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Reltter, Marc
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
; FILE REFERENCE: 014058-013520US
; CURRENT APPLICATION NUMBER: US/10/040,862
; PRIOR FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: US 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/223,378
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: US 09/796,692
; PRIOR FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 10467
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9051
; LENGTH: 390
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-040-862-9051

Query Match      29.5%; Score 34.2; DB 14; Length 390;
```



Best Local Similarity 58.3%; Pred. No. 0.062;  
Matches 60; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 10 ACTGCTGGAATACCTTTGTAGAAAACCAAGAAAGAACTTTCAAGCTGGAGAGGCTGC 69  
Db 165 ACTGCTGGAGACACCTTTGTAGACCAACAGGATGTCCTTCCAGCCCTGGATGACTAG 224

QY 70 ATGAAATTCAGTTCCTCTCCAGACAGCTTGGCGCATCT 112  
Db 225 ATGAGCAGACCAAGACCTGAGTGGAGGCTGGCGCATCT 267

## RESULT 8

US-10-057-475B-9051  
; Sequence 9051, Application US/10057475B  
; Publication No. US20040002068A1  
; GENERAL INFORMATION:  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Mannon, Jane  
; APPLICANT: Clapper, Jonathan David  
; APPLICANT: Wang, Aijun  
; APPLICANT: Ordenez, Nadia  
; APPLICANT: Carter, Lauren  
; APPLICANT: McNeill, Patricia Diane  
; APPLICANT: Corixa Corporation  
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy  
; FILE REFERENCE: 014058-014402US  
; CURRENT APPLICATION NUMBER: US/10/057,475B  
; CURRENT FILING DATE: 2002-01-22  
; PRIOR APPLICATION NUMBER: US 60/186,126  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: US 60/190,479  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: US 60/200,545  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: US 60/200,303  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 60/200,779  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 60/200,999  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: US 60/202,084  
; PRIOR FILING DATE: 2000-05-04  
; PRIOR APPLICATION NUMBER: US 60/206,201  
; PRIOR FILING DATE: 2000-05-22  
; PRIOR APPLICATION NUMBER: US 60/218,950  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: US 60/222,903  
; PRIOR FILING DATE: 2000-08-03  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 10979  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 9051  
; LENGTH: 390  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-057-475B-9051

Query Match 29.5%; Score 34.2; DB 15; Length 390;  
Best Local Similarity 58.3%; Pred. No. 0.062;  
Matches 60; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 10 ACTGCTGGAATACCTTTGTAGAAAACCAAGAAAGAACTTTCAAGCTGGAGAGGCTGC 69  
Db 165 ACTGCTGGAGACACCTTTGTAGACCAACAGGATGTCCTTCCAGCCCTGGATGACTAG 224

QY 70 ATGAAATTCAGTTCCTCTCCAGACAGCTTGGCGCATCT 112  
Db 225 ATGAGCAGACCAAGACCTGAGTGGAGGCTGGCGCATCT 267

RESULT 9  
US-10-154-884B-9051  
; Sequence 9051, Application US/10154884B  
; Publication No. US20040005561A1  
; GENERAL INFORMATION:  
; APPLICANT: Gaiger, Alexander  
; APPLICANT: Algate, Paul A.  
; APPLICANT: Mannon, Jane  
; APPLICANT: Reiter, Marc W.  
; APPLICANT: Corixa Corporation  
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy  
; FILE REFERENCE: 014058-013521US  
; CURRENT APPLICATION NUMBER: US/10/154,884B  
; CURRENT FILING DATE: 2002-05-23  
; PRIOR APPLICATION NUMBER: US 60/186,126  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: US 60/190,479  
; PRIOR FILING DATE: 2000-03-17  
; PRIOR APPLICATION NUMBER: US 60/200,545  
; PRIOR FILING DATE: 2000-04-27  
; PRIOR APPLICATION NUMBER: US 60/200,303  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 60/200,779  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: US 60/200,999  
; PRIOR FILING DATE: 2000-05-01  
; PRIOR APPLICATION NUMBER: US 60/202,084  
; PRIOR FILING DATE: 2000-05-04  
; PRIOR APPLICATION NUMBER: US 60/206,201  
; PRIOR FILING DATE: 2000-05-22  
; PRIOR APPLICATION NUMBER: US 60/218,950  
; PRIOR FILING DATE: 2000-07-14  
; PRIOR APPLICATION NUMBER: US 60/222,903  
; PRIOR FILING DATE: 2000-08-03  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 11290  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 9051  
; LENGTH: 390  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-154-884B-9051

Query Match 29.5%; Score 34.2; DB 15; Length 390;  
Best Local Similarity 58.3%; Pred. No. 0.062;  
Matches 60; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 10 ACTGCTGGAATACCTTTGTAGAAAACCAAGAAAGAACTTTCAAGCTGGAGAGGCTGC 69  
Db 165 ACTGCTGGAGACACCTTTGTAGACCAACAGGATGTCCTTCCAGCCCTGGATGACTAG 224

QY 70 ATGAAATTCAGTTCCTCTCCAGACAGCTTGGCGCATCT 112  
Db 225 ATGAGCAGACCAAGACCTGAGTGGAGGCTGGCGCATCT 267

RESULT 10  
US-10-460-923-1  
; Sequence 1, Application US/10460923  
; Publication No. US20040009951A1  
; GENERAL INFORMATION:  
; APPLICANT: MALIM, Michael H.  
; APPLICANT: SHEEHY, Ann M.  
; APPLICANT: HARRIS, Reuben S.  
; APPLICANT: BISHOP, Kate N.  
; APPLICANT: NEUBERGER, Michael S.  
; APPLICANT: GADDIS, Nathan C.  
; APPLICANT: SIMON, James H.M.  
; TITLE OF INVENTION: DNA Desamination Mediates Innate Immunity to Retroviral Infection  
; FILE REFERENCE: 22253-74380  
; CURRENT APPLICATION NUMBER: US/10/460,923  
; CURRENT FILING DATE: 2003-06-13



```
; PRIOR APPLICATION NUMBER: US 60/388,513
; PRIOR FILING DATE: 2002-06-13
; PRIOR APPLICATION NUMBER: US 60/472,952
; PRIOR FILING DATE: 2003-05-23
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1155
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-460-923-1

Query Match      29.5%; Score 34.2; DB 15; Length 1155;
Best Local Similarity 58.3%; Pred. No. 0.099;
Matches 60; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 10 ACTGCTGGAATCTTTGTAGAAAACCAAGAACTTCAAGCTGGAGGGCTGC 69
Db 1034 ACTGCTGGGACACCTTTGTGACACCAAGGATGTCCCTTCAGCCCTGGATGACTAG 1093
QY 70 ATGAAATTCAGTTCCTCTCCAGACAGCTTCGGGCGCATCTT 112
Db 1094 ATGAGCAGCAGCAAGACCTAGTGGGAGGCTGGGGCGCATCTT 1136

RESULT 11
US-09-880-107-3282
; Sequence 3282, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Scherf, Uwe
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-MO
; CURRENT APPLICATION NUMBER: US/09/880,107
; PRIOR FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3282
; LENGTH: 1348
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 U03891
; NAME/KEY: unsure
; LOCATION: (1)..(1348)
; OTHER INFORMATION: n = a or c or g or t
US-09-880-107-3282

Query Match      29.5%; Score 34.2; DB 9; Length 1348;
Best Local Similarity 58.3%; Pred. No. 0.11;
Matches 60; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 10 ACTGCTGGAATCTTTGTAGAAAACCAAGAACTTCAAGCTGGAGGGCTGC 69
Db 554 ACTGCTGGGACACCTTTGTGACACCAAGGATGTCCCTTCAGCCCTGGATGACTAG 613
QY 70 ATGAAATTCAGTTCCTCTCCAGACAGCTTCGGGCGCATCTT 112
Db 614 ATGAGCAGCAGCAAGCCCTAGTGGGAGGCTGGGGCGCATCTT 656

RESULT 12
US-09-729-674-173
; Sequence 173, Application US/09729674
; Patent No. US2001003935A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: Lavalley, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Metberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Agostino, Michael J.
; APPLICANT: Steininger II, Robert J.
; APPLICANT: Spaulding, Vikki
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Pechtel, Kim
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYPEPTIDES ENCODING THEM
; FILE REFERENCE: 6055-64X
; CURRENT APPLICATION NUMBER: US/09/729,674
; PRIOR FILING DATE: 2000-12-04
; PRIOR APPLICATION NUMBER: 09/539,330
; PRIOR FILING DATE: 2000-03-30
; NUMBER OF SEQ ID NOS: 283
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 173
; LENGTH: 1534
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-729-674-173

Query Match      29.5%; Score 34.2; DB 12; Length 1534;
Best Local Similarity 58.3%; Pred. No. 0.11;
Matches 60; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 10 ACTGCTGGAATCTTTGTAGAAAACCAAGAACTTCAAGCTGGAGGGCTGC 69
Db 1084 ACTGCTGGGACACCTTTGTGACACCAAGGATGTCCCTTCAGCCCTGGATGACTAG 1143
QY 70 ATGAAATTCAGTTCCTCTCCAGACAGCTTCGGGCGCATCTT 112
Db 1144 ATGAGCAGCAGCAAGACCTAGTGGGAGGCTGGGGCGCATCTT 1186

RESULT 13
US-10-240-425-1099
; Sequence 1099, Application US/10240425
; Publication No. US20040033502A1
; GENERAL INFORMATION:
; APPLICANT: Williams, Amanda
; APPLICANT: Boland, Joseph F.
; APPLICANT: Lord, Reginald V.
; APPLICANT: Alvarez, Chris
; APPLICANT: Wetzel, Jon C.
; APPLICANT: Scherf, Uwe
; APPLICANT: Vockley, Joseph G.
; TITLE OF INVENTION: Gene Expression Profiles in Esophageal Tissue
; FILE REFERENCE: 44921-5026
; CURRENT APPLICATION NUMBER: US/10/240,425
; PRIOR FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: PCT/US01/09847
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: US 60/193,446
; PRIOR FILING DATE: 2000-03-31
; NUMBER OF SEQ ID NOS: 1588
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1099
; LENGTH: 201143
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20040033502A1 AL022318
US-10-240-425-1099

Query Match      29.5%; Score 34.2; DB 12; Length 201143;
Best Local Similarity 58.3%; Pred. No. 0.93;
```



Matches 60; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 10 ACTGCTGGAATACCTTTGTAGAAAACCGAAGAACTTCAAGCTGGAGGGCTGC 69

Db 77914 ACTGCTGGACACCTTTGTGACACCAAGATGTCCCTTCAGCCCTGGATGACTG 77973

QY 70 ATGAAATTCAGTTCGTCTCTCCAGACAGCTTGGGCGCATCCT 112

Db 77974 ATGAGCACAGCCCAAGCCCTGATGGAGGCTGGCGGCGCATTCCT 78016

## RESULT 14

US-10-424-599-23145

/ Sequence 23145; Application US/10424599

/ Publication No. US20040031072A1

/ GENERAL INFORMATION:

/ APPLICANT: La Rosa Thomas J

/ APPLICANT: Kovalic David K

/ APPLICANT: Zhou Yinhua

/ APPLICANT: Cao Yongwei

/ TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With

/ TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

/ FILE REFERENCE: 38-21(53223)B

/ CURRENT APPLICATION NUMBER: US/10/424,599

/ NUMBER OF SEQ ID NOS: 2003-04-28

/ SEQ ID NO 23145

/ LENGTH: 274

/ TYPE: DNA

/ ORGANISM: Glycine max

/ FEATURE:

/ OTHER INFORMATION: Clone ID: PAT\_MRT3847\_120901C.1

US-10-424-599-23145

## Query Match

Best Local Similarity 26.4%; Score 30.6; DB 12; Length 274;

Matches 54; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 23 TTTTGTAGAAAACCGAAGAACTTCAAGCTTGGAGGGCTGCATGAAATTCAGT 82

Db 99 TTATGACCCAACTCAAGAGATCTTAGAGACTGCTGAGGCTCACATTTATTTAGT 158

QY 83 TCGTCTCTCCAGACAGCTTGGGCGCATCCTTT 115

Db 159 CTGCTCCAAAGCATGTGTAATCTCTTT 191

## RESULT 15

US-09-796-692-5737

/ Sequence 5737; Application US/09796692

/ Publication No. US2002019362A1

/ GENERAL INFORMATION:

/ APPLICANT: Gaiger, Alexander

/ APPLICANT: Algate, Paul A.

/ APPLICANT: Mannion, Jane

/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY

/ TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES

/ FILE REFERENCE: 2077.001200

/ CURRENT APPLICATION NUMBER: US/09/796,692

/ PRIOR FILING DATE: 2001-03-01

/ PRIOR APPLICATION NUMBER: 60/186,126

/ PRIOR FILING DATE: 2000-03-01

/ PRIOR APPLICATION NUMBER: 60/190,479

/ PRIOR FILING DATE: 2000-03-17

/ PRIOR APPLICATION NUMBER: 60/200,545

/ PRIOR FILING DATE: 2000-04-27

/ PRIOR APPLICATION NUMBER: 60/200,303

/ PRIOR FILING DATE: 2000-04-28

/ PRIOR APPLICATION NUMBER: 60/200,779

/ PRIOR FILING DATE: 2000-04-28

/ PRIOR APPLICATION NUMBER: 60/200,999

/ PRIOR FILING DATE: 2000-05-01

/ PRIOR APPLICATION NUMBER: 60/202,084

/ PRIOR FILING DATE: 2000-05-04

/ PRIOR APPLICATION NUMBER: 60/206,201

/ PRIOR FILING DATE: 2000-05-22

/ PRIOR APPLICATION NUMBER: 60/218,950

/ PRIOR FILING DATE: 2000-07-14

/ PRIOR APPLICATION NUMBER: 60/222,903

/ PRIOR FILING DATE: 2000-08-03

/ PRIOR APPLICATION NUMBER: 60/223,416

/ PRIOR FILING DATE: 2000-08-04

/ PRIOR APPLICATION NUMBER: 60/223,378

/ PRIOR FILING DATE: 2000-08-07

/ NUMBER OF SEQ ID NOS: 9597

/ SOFTWARE: FastSeq for Windows Version 3.0

/ SEQ ID NO 5737

/ LENGTH: 328

/ TYPE: DNA

/ ORGANISM: Homo sapiens

US-09-796-692-5737

## Query Match

Best Local Similarity 26.2%; Score 30.4; DB 9; Length 328;

Matches 61; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 1 ATATTTTCTGCTGGAATACCTTTGTAGAAAACCGAAGAACTTCAAGCTGGG 60

Db 130 ATTTGTATCTTGTGAAAACCTTGTGTACGTGATGATGAGCCATTCAAGCTTGA 189

QY 61 AAGGCTGCATGAAAATTCAGTTCGTCTCCAGACAGCTTGGGCGCATCCT 112

Db 190 AGGACTACAAACCACTTTCGACTTCTGAAAAGAGGCTACGGAGATTCT 241

Search completed: March 13, 2004, 05:31:10  
Job time : 44.7067 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 12, 2004, 13:19:34 ; Search time 369.373 Seconds  
(without alignments)  
9378.080 Million cell updates/sec

Title: US-09-966-880A-14

Perfect score: 116  
Sequence: 1 atattttactgctggaat.....agcttcgcccacattttg 116

Scoring table: IDENTITY NJC  
Gapop 10\_0, Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estda:\*  
2: em\_estdm:\*  
3: em\_estln:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_liv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vit:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vit:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	114.4	98.6	693	10	BF975096 602245679
2	114.4	98.6	743	12	BG686133 602638412
3	114.4	98.6	820	12	BG757089 602715124
4	114.4	98.6	872	12	BG758510 602712721

## ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	TITLE	AUTHORS	JOURNAL	COMMENT
BF975096	602245679	NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4336722 5',	BF975096	BF975096	GI:12342311	Homo sapiens (human)	Homo sapiens	1 (bases 1 to 693)	NIH-MGC http://mgc.nci.nih.gov/	Unpublished (1999)	National Institutes of Health, Mammalian Gene Collection (MGC)	Contact: Robert Strussberg, Ph.D. Email: cga@bbs-riemail.nih.gov Tissue Procurement: Louis M. Staudt, M.D., Ph.D. cDNA Library Preparation: Ling Hong/Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov Plate: LNCM1209 row: n column: 19 High quality sequence stop: 692.
5	114.4	98.6	953	13	BQ055440							BQ055440 AGENCOURT
6	114.4	98.6	1052	13	BQ055935							BQ055935 AGENCOURT
7	114.4	98.6	1201	9	AL559877							AL559877 AL559877
8	112.8	97.2	693	12	BG757392							BG757392 602711022
9	112.8	97.2	853	13	BX464579							BX464579 BX464579
10	112.8	97.2	942	10	BF975166							BF975166 602244657
11	103.4	89.1	1201	13	BX402063							BX402063 BX402063
12	100	86.2	1034	13	BG755005							BG755005 602711511
13	98.4	84.8	442	28	BH302559							BH302559 CH230-100
14	96.8	83.4	522	12	BG144705							BG144705 UT73107.Y
15	96.8	83.4	1292	11	AK080144							AK080144 Mus muscu
16	90.6	78.1	843	12	BG758815							BG758815 602713177
17	84.2	72.6	920	13	BX392040							BX392040 BX392040
18	81.6	70.3	653	10	BB637360							BB637360 BB637360
19	76	65.5	757	9	AJ446140							AJ446140 AJ446140
20	74.4	64.1	696	9	AJ453647							AJ453647 AJ453647
21	72.2	62.2	388	14	CB171719							CB171719 B21603020
22	68.2	58.8	688	9	AJ450317							AJ450317 AJ450317
23	60.8	52.4	729	9	AJ450296							AJ450296 AJ450296
24	54.6	47.1	650	9	AJ449745							AJ449745 AJ449745
25	46.4	40.0	498	29	DR38019T							DR38019T Danio rer
26	45.4	39.1	644	14	CD469636							CD469636 LeukOS2_7
27	45.4	39.1	707	14	CD470713							CD470713 LeukOS4_6
28	43.8	37.8	642	14	CD470941							CD470941 LeukOS5_3
29	43.8	37.8	682	14	CD471110							CD471110 LeukOS5_4
30	42.2	36.4	597	14	CD470940							CD470940 LeukOS5_3
31	42.2	36.4	632	14	CD471262							CD471262 LeukOS5_1
32	42.2	36.4	653	14	CD528537							CD528537 LeukON3_2
33	42.2	36.4	670	14	CD470816							CD470816 LeukOS5_2
34	42.2	36.4	690	14	CD535895							CD535895 LeukON5_5
35	42.2	36.4	731	14	CD469635							CD469635 LeukOS2_7
36	42.2	36.4	740	14	CD469716							CD469716 LeukOS2_8
37	41.2	35.5	690	14	CD471162							CD471162 LeukOS5_4
38	40.6	35.0	630	14	CD469721							CD469721 LeukOS2_8
39	40.6	35.0	637	14	CD470714							CD470714 LeukOS4_6
40	40.6	35.0	640	14	CD470661							CD470661 LeukOS4_6
41	40.6	35.0	643	14	CD465467							CD465467 LeukON1_4
42	40.4	34.8	197	14	CB286686							CB286686 CMD4_4 B05
43	40.4	34.8	197	14	CF722800							CF722800 D12 LXF00
44	37	31.9	541	10	BF238155							BF238155 601811880
45	36.4	31.4	647	14	CD469225							CD469225 LeukOS2_2



# FEATURES

Location/Qualifiers

1. 693  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4336722"  
/issue\_type="primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: B-cells; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 98.6%; Score 114.4; DB 10; Length 693;  
Best Local Similarity 99.1%; Pred. No. 1.5e-26;  
Matches 115; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATTATTTTACTGCTGGAATCTTTGTAGAAAACGAGAAAGCTTCAAGCCTGGG 60  
DB 204 ATTATTTTACTGCTGGAATCTTTGTAGAAAACGAGAAAGCTTCAAGCCTGGG 263

QY 61 AAGGCTGCATGAAATTCAGTTCCTCCAGACAGCTTCGGCGCATCCTTTTG 116  
DB 264 AAGGCTGCATGAAATTCAGTTCCTCCAGACAGCTTCGGCGCATCCTTTTG 319

RESULT 2 743 bp mRNA linear EST 01-MAY-2001  
BG686133 602638412F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4766234 5',  
LOCUS mRNA sequence.

ACCESSION BG686133.1 GI:13917530  
VERSION EST.  
KEYWORDS Homo sapiens (human)  
SOURCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
ORGANISM Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 743)  
AUTHORS NIH-MGC http://mgs.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.lnl.gov  
Plate: LNCM626 row: 9 column: 03  
High quality sequence stop: 740.

FEATURES  
source location/Qualifiers

1. 743  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4766234"  
/issue\_type="primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_id="NIH\_MGC\_48"  
/note="Organ: B-cells; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp

## ORIGIN

Query Match 98.6%; Score 114.4; DB 12; Length 743;  
Best Local Similarity 99.1%; Pred. No. 1.5e-26;  
Matches 115; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATTATTTTACTGCTGGAATCTTTGTAGAAAACGAGAAAGCTTCAAGCCTGGG 60  
DB 505 ATTATTTTACTGCTGGAATCTTTGTAGAAAACGAGAAAGCTTCAAGCCTGGG 564

QY 61 AAGGCTGCATGAAATTCAGTTCCTCCAGACAGCTTCGGCGCATCCTTTTG 116  
DB 565 AAGGCTGCATGAAATTCAGTTCCTCCAGACAGCTTCGGCGCATCCTTTTG 620

for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 98.6%; Score 114.4; DB 12; Length 820;  
Best Local Similarity 99.1%; Pred. No. 1.5e-26;  
Matches 115; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATTATTTTACTGCTGGAATCTTTGTAGAAAACGAGAAAGCTTCAAGCCTGGG 60  
DB 505 ATTATTTTACTGCTGGAATCTTTGTAGAAAACGAGAAAGCTTCAAGCCTGGG 564

QY 61 AAGGCTGCATGAAATTCAGTTCCTCCAGACAGCTTCGGCGCATCCTTTTG 116  
DB 565 AAGGCTGCATGAAATTCAGTTCCTCCAGACAGCTTCGGCGCATCCTTTTG 620

RESULT 3 820 bp mRNA linear EST 15-MAY-2001  
BG757089 602715124F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4855517 5',  
LOCUS mRNA sequence.

ACCESSION BG757089  
VERSION BG757089.1 GI:14067742  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 820)  
AUTHORS NIH-MGC http://mgs.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.lnl.gov  
Plate: LNCM1704 row: 0 column: 06  
High quality sequence stop: 675.

FEATURES  
source location/Qualifiers

1. 820  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4855517"  
/issue\_type="primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_id="NIH\_MGC\_48"  
/note="Organ: B-cells; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 98.6%; Score 114.4; DB 12; Length 820;  
Best Local Similarity 99.1%; Pred. No. 1.5e-26;  
Matches 115; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATTATTTTACTGCTGGAATCTTTGTAGAAAACGAGAAAGCTTCAAGCCTGGG 60



Db 512 ATTATTTTACTGCTGGAATACCTTTGTAGAAAACCATGAAGAACTTCAAGCTGGG 571

QY 61 AAGGGCTGCATGAATAATTCAGTTGCTCTCCAGACAGCTTGGCGCATCTTTTG 116

Db 572 AAGGGCTGCATGAATAATTCAGTTGCTCTCCAGACAGCTTGGCGCATCTTTTG 627

RESULT 4

LOCUS BG758510

DEFINITION 872 bp mRNA linear EST 15-MAY-2001

60271272.F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4853069.5', mRNA sequence.

ACCESSION BG758510

VERSION BG758510

KEYWORDS GI:114069163

SOURCE EST.

ORGANISM Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 872)

NIH-MGC <http://mgs.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)

Tissue Procurement: Louis M. Staudt, M.D., Ph.D.

CDNA Library Preparation: Ling Hong/Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Plate: L1CM2108 row: 1 column: 06

High quality sequence stop: 836.

Location/Qualifiers

1. 872

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:4853069"

/tissue\_type="primary B-cells from tonsils (cell line)"

/lab\_host="DH10B (phage-resistant)"

/clone\_lib="NIH\_MGC\_48"

/note="Organ: B-cells; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

ORIGIN

Query Match 98.6%; Score 114.4; DB 12; Length 872;

Best Local Similarity 99.1%; Pred. No. 1.5e-26;

Matches 115; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATTATTTTACTGCTGGAATACCTTTGTAGAAAACCATGAAGAACTTCAAGCTGGG 60

Db 493 ATTATTTTACTGCTGGAATACCTTTGTAGAAAACCATGAAGAACTTCAAGCTGGG 552

QY 61 AAGGGCTGCATGAATAATTCAGTTGCTCTCCAGACAGCTTGGCGCATCTTTTG 116

Db 553 AAGGGCTGCATGAATAATTCAGTTGCTCTCCAGACAGCTTGGCGCATCTTTTG 608

RESULT 5

LOCUS BQ065440

DEFINITION 953 bp mRNA linear EST 02-APR-2002

6855061 NIH\_MGC\_99 Homo sapiens cDNA clone IMAGE:5929977

5', mRNA sequence.

ACCESSION BQ065440

VERSION BQ065440.1 GI:19894486

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 953)

NIH-MGC <http://mgs.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)

Tissue Procurement: Lou Staudt

CDNA Library Preparation: Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Plate: L1CM2108 row: p column: 10

High quality sequence stop: 634.

Location/Qualifiers

1. 953

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="IMAGE:5929977"

/tissue\_type="lymphoma, cell line"

/lab\_host="DH10B (phage-resistant)"

/clone\_lib="NIH\_MGC\_99"

/note="Organ: lymph; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

ORIGIN

Query Match 98.6%; Score 114.4; DB 13; Length 953;

Best Local Similarity 99.1%; Pred. No. 1.6e-26;

Matches 115; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATTATTTTACTGCTGGAATACCTTTGTAGAAAACCATGAAGAACTTCAAGCTGGG 60

Db 491 ATTATTTTACTGCTGGAATACCTTTGTAGAAAACCATGAAGAACTTCAAGCTGGG 550

QY 61 AAGGGCTGCATGAATAATTCAGTTGCTCTCCAGACAGCTTGGCGCATCTTTTG 116

Db 551 AAGGGCTGCATGAATAATTCAGTTGCTCTCCAGACAGCTTGGCGCATCTTTTG 606

RESULT 6

LOCUS BQ055935

DEFINITION 1052 bp mRNA linear EST 29-MAR-2002

6796291 NIH\_MGC\_99 Homo sapiens cDNA clone IMAGE:5808181

5', mRNA sequence.

ACCESSION BQ055935

VERSION BQ055935

KEYWORDS GI:19815262

SOURCE EST.

ORGANISM Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1052)

NIH-MGC <http://mgs.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)

Tissue Procurement: Lou Staudt



CDNA Library Preparation: Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.lnl.gov>  
 Plate: LNCM2051 row: m column: 14  
 High quality sequence stop: 665.  
 Location/Qualifiers  
 1..1052

## FEATURES

source

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5808181"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_99"
/note="Organ: lymph; Vector: pOTB7, Site_1: XhoI, Site_2:
EcoRI; CDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."

```

## ORIGIN

Query Match 98.6%; Score 114.4; DB 13; Length 1052;  
 Best Local Similarity 99.1%; Pred. No. 1.6e-26;  
 Matches 115; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATTATTTTACTGCTGGAATCTTTGTAGAAAACCGAAAGACTTTCAAGCCTGGG 60  
 |||  
 DB 491 ATTATTTTACTGCTGGAATCTTTGTAGAAAACCGAAAGACTTTCAAGCCTGGG 550  
 |||  
 QY 61 AAGGCGTCAGTAATAATTCAGTTCCTCCAGACAGCTTCGGCGCATCCTTTTG 116  
 |||  
 DB 551 AAGGCGTCAGTAATAATTCAGTTCCTCCAGACAGCTTCGGCGCATCCTTTTG 606  
 |||

RESULT 7 1201 bp mRNA linear EST 31-MAY-2003  
 AL559877  
 LOCUS  
 DEFINITION AL559877 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens CDNA  
 clone CSODG003B14 5-PRIME, mRNA sequence.  
 ACCESSION AL559877  
 VERSION AL559877.2 GI:31284008  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 1201)  
 REFERENCE  
 AUTHORS Li, M.B., Gruber, C., Jesse, J. and Polayes, D.  
 TITLE Full-length CDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT On Feb 15, 2001 this sequence version replaced gi:12905793.  
 Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 104 91006 EVRY cedex - France  
 Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 6672.r for  
 more information about this cluster, see  
[http://www.genoscope.cns.fr/](http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CSODG003DA07QPL&cluster=6672.r)  
<http://fulllength.invitrogen.com/InvitrogenCorporation1600>  
 Paraday Avenue Genoscope sequence ID: CSODG003DA07QPL.  
 Location/Qualifiers  
 1..1201  
 /organism="Homo sapiens"

## FEATURES

source

```

1..1201
/organism="Homo sapiens"

```

```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODG003B14"
/tissue_type="B CELLS (RAMOS CELL LINE)"
/clone_lib="RAMOS CELL LINE"
/note="Organ: B-cells; Vector: pOTB7, Site_1: XhoI,
Site_2: EcoRI; CDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAGCAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

```

## ORIGIN

Query Match 98.6%; Score 114.4; DB 9; Length 1201;  
 Best Local Similarity 99.1%; Pred. No. 1.6e-26;  
 Matches 115; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATTATTTTACTGCTGGAATCTTTGTAGAAAACCGAAAGACTTTCAAGCCTGGG 60  
 |||  
 DB 519 ATTATTTTACTGCTGGAATCTTTGTAGAAAACCGAAAGACTTTCAAGCCTGGG 578  
 |||  
 QY 61 AAGGCGTCAGTAATAATTCAGTTCCTCCAGACAGCTTCGGCGCATCCTTTTG 116  
 |||  
 DB 579 AAGGCGTCAGTAATAATTCAGTTCCTCCAGACAGCTTCGGCGCATCCTTTTG 634  
 |||

RESULT 8 693 bp mRNA linear EST 15-MAY-2001  
 BG757392  
 LOCUS  
 DEFINITION BG757392 NIH\_MGC\_48 Homo sapiens CDNA clone IMAGE:4851580 5',  
 mRNA sequence.  
 ACCESSION BG757392  
 VERSION BG757392.1 GI:14068045  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 693)  
 REFERENCE  
 AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgabs-remail.nih.gov  
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
 CDNA Library Preparation: Ling Hong/Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.lnl.gov>  
 Plate: LNCM1694 row: k column: 05  
 High quality sequence stop: 693.  
 Location/Qualifiers  
 1..693  
 /organism="Homo sapiens"
 /mol\_type="mRNA"
 /db\_xref="taxon:9606"
 /clone="IMAGE:4851580"
 /tissue\_type="primary B-cells from tonsils (cell line)"
 /lab\_host="DH10B (phage-resistant)"
 /clone\_lib="NIH\_MGC\_48"
 /note="Organ: B-cells; Vector: pOTB7, Site\_1: XhoI,
 Site\_2: EcoRI; CDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCAGCAG(G). Size-selected >500bp
 for average insert size 1.8kb. Library constructed by Ling
 Hong in the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH\_MGC Library."

## FEATURES

source

## ORIGIN



Query Match 97.2%; Score 112.8; DB 12; Length 693;  
 Best Local Similarity 98.3%; Pred. No. 5.1e-26;  
 Matches 114; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATATTTTACTGCTGGAATACCTTTGTAGAAAAACCAAGAAAGACTTTCAAGCCTGGG 60  
 DB 509 ATATTTTACTGCTGGAATACCTTTGTAGAAAAACCAAGAAAGACTTTCAAGCCTGGG 568

QY 61 AAGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGGCGCATCCTTTTG 116  
 DB 569 AAGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGGCGCATCCTTTTG 624

RESULT 9  
 EX464579 853 bp mRNA linear EST 22-MAY-2003  
 LOCUS BX464579  
 DEFINITION BX464579 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens CDNA clone CS0D003YB14 5-PRIME, mRNA sequence.  
 ACCESSION BX464579  
 VERSION BX464579.1 GI:31031641  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
 TITLE Full-length CDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 6672.r For more information about this cluster, see  
 http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CS1D001ZF10QPLcluster=6672.r. Contact :  
 Feng Liang Email: fliang@lifetech.com URL :  
 http://fulllength.invitrogen.com/invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID : CS1D001ZF10QPL.  
 Location/Qualifiers  
 1..853  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0D003YB14"  
 /issue\_type="B CELLS (RAMOS CELL LINE)"  
 /cell\_line="RAMOS CELL LINE"  
 /clone\_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"  
 /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

ORIGIN

Query Match 97.2%; Score 112.8; DB 13; Length 853;  
 Best Local Similarity 98.3%; Pred. No. 5.1e-26;  
 Matches 114; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATATTTTACTGCTGGAATACCTTTGTAGAAAAACCAAGAAAGACTTTCAAGCCTGGG 60  
 DB 513 ATATTTTACTGCTGGAATACCTTTGTAGAAAAACCAAGAAAGACTTTCAAGCCTGGG 572

QY 61 AAGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGGCGCATCCTTTTG 116  
 DB 573 AAGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGGCGCATCCTTTTG 628

RESULT 10  
 BF975166 942 bp mRNA linear EST 22-JAN-2001  
 LOCUS BF975166

DEFINITION 602244657r1 NIH\_MGC\_48 Homo sapiens CDNA clone IMAGE:4335639 5', mRNA sequence.  
 ACCESSION BF975166  
 VERSION BF975166.1 GI:12342381  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE NIH-MGC http://nigc.nci.nih.gov/  
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
 TITLE Unpublished (1999)  
 JOURNAL Contact: Robert Strausberg, Ph.D.  
 COMMENT Email: cga@bbs-remail.nih.gov  
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: L10C1207 row: a column: 16  
 High quality sequence stop: 707.  
 Location/Qualifiers  
 1..942  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4335639"  
 /issue\_type="Primary B-cells from tonsils (cell line)"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_48"  
 /note="Organ: B-cells; Vector: pOTB7; Site: 1; XhoI; Site: 2; EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAAGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald W. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

ORIGIN

Query Match 97.2%; Score 112.8; DB 10; Length 942;  
 Best Local Similarity 98.3%; Pred. No. 5.2e-26;  
 Matches 114; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATATTTTACTGCTGGAATACCTTTGTAGAAAAACCAAGAAAGACTTTCAAGCCTGGG 60  
 DB 506 ATATTTTACTGCTGGAATACCTTTGTAGAAAAACCAAGAAAGACTTTCAAGCCTGGG 565

QY 61 AAGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGGCGCATCCTTTTG 116  
 DB 566 AAGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGGCGCATCCTTTTG 621

RESULT 11  
 BX402063 1201 bp mRNA linear EST 13-MAY-2003  
 LOCUS BX402063  
 DEFINITION BX402063 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED  
 ACCESSION BX402063  
 VERSION BX402063.1 GI:30626645  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
 TITLE Full-length CDNA libraries and normalization  
 JOURNAL Unpublished (2001)



## COMMENT

Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 6672.r For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0D012DB09QPI&cluster=6672.r. Contact :  
Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0D012DB09QPI.

## FEATURES

## SOURCE

```
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D012YD18"
/cell_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"
/cell_line="RAMOS CELL LINE"
/clone_id="Homo sapiens B CELLS (RAMOS CELL LINE) COT
25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo (dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
```

## ORIGIN

## Query Match

Best Local Similarity 89.1%; Score 103.4; DB 13; Length 1201;  
Matches 115; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

```
QY 1 ATTATTTTACTGCTGGAAATCT-TTTGTAGAAACACGAAAGACTTTCAAGCCTGG 59
DB 579 ATTATTTTACTGCTGGAAATCTTTGTAGAAACACGAAAGACTTTCAAGCCTGG 638
DB 639 GAAGGCTGATGAATAATTCAGTTCCTCTCCAGACAGCTTGGGCGATCCTTTTG 695
```

## RESULT 12

## BG755005

LOCUS 1034 bp mRNA linear EST 15-MAY-2001

## DEFINITION

60271151F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4851819 5',

## ACCESSION

BG755005 mRNA sequence.

## VERSION

BG755005.1 GI:14065658

## KEYWORDS

EST.

## SOURCE

Homo sapiens (human)

## ORGANISM

Homo sapiens

## REFERENCE

1 (bases 1 to 1034)

## AUTHORS

NH-MGC http://mgs.nci.nih.gov/.

## TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

## JOURNAL

Unpublished (1999)

## COMMENT

Contact: Robert Strausberg, Ph.D.  
Email: cgabs@remail.nih.gov  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1CM1695 row: e column: 04  
High quality sequence stop: 622.  
Location/Qualifiers

## FEATURES

## source

```
1..1034
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
```

## ORIGIN

Query Match 86.2%; Score 100; DB 12; Length 1034;  
Best Local Similarity 91.4%; Pred. No. 8e-22; Indels 0; Gaps 0;  
Matches 106; Conservative 0; Mismatches 10;

```
QY 1 ATTATTTTACTGCTGGAAATCTTTGTAGAAACACGAAAGACTTTCAAGCCTGG 60
DB 45 ATTATTTTACTGCTGGAAATCTTTGTAGAAACACGAAAGACTTTCAAGCCTGG 104
QY 61 AAGGCTGATGAATAATTCAGTTCCTCTCCAGACAGCTTGGGCGATCCTTTTG 116
DB 105 AAGGCTGATGAATAATTCAGTTCCTCTCCAGACAGCTTGGGCGATCCTTTTG 160
```

## RESULT 13

## BH302559/c

## LOCUS

BH302559 CH230-100C8.TV CHORI-230 Segment 1 Rattus norvegicus genomic clone

## DEFINITION

CH230-100C8, genomic survey sequence.

## ACCESSION

BH302559

## VERSION

BH302559.1 GI:17214967

## KEYWORDS

GSS.

## SOURCE

Rattus norvegicus (Norway rat)

## ORGANISM

Rattus norvegicus

## REFERENCE

1 (bases 1 to 442)

## AUTHORS

Zhao,S., Shetty,J., Shatman,S., Teagay,G., Geer,K.,

Shvartsbeyn,A., Gebregorgis,E., Overton,L., Russell,D., Chen,D.,

Riggs,F., de Jong,P. and Fraser,C.M.

Rat BAC End Sequences from Library CHORI-230 EcORI segment

Unpublished (1999)

Other GSSs: CH230-100C8.TV

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the rat BAC library CHORI-230

(http://www.chori.org/bacpac/rat230.htm). For BAC library

availability, please contact Pieter de Jong (pdjong@mail.choi.org).

Clones may be purchased from BACPAC Resources

(http://www.chori.org/bacpac/or ordering information.htm). BAC end

page: http://www.tigr.org/tcb/bac\_end/rat/bac\_end\_intro.html

Plate: 100 row: C column: 8

Seq primer: T7

Class: BAC ends.

Location/Qualifiers

```
1..442
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="BN/SSNHsd/MCW"
/db_xref="taxon:10116"
/clone="CH230-100C8"
/sex="Female"
/cell_type="Brain"
```







## COMMENT

Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,  
 URL: <http://genome.gsc.riken.go.jp/>, Tel:81-45-503-9222,  
 Fax:81-45-503-9216)  
 cDNA library was prepared and sequenced in Mouse Genome  
 Encyclopedia Project of Genome Exploration Research Group in Riken  
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
 Division of Experimental Animal Research in Riken contributed to  
 prepare mouse tissues.  
 Please visit our web site for further details.  
 URL: <http://genome.gsc.riken.go.jp/>  
 URL: <http://fantom.gsc.riken.go.jp/>.

## FEATURES

## source

Location/Qualifiers  
 1..1292  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57Bl/6J"  
 /db\_xref="FANTOM\_DB:A530070C03"  
 /db\_xref="MGI:2405961"  
 /db\_xref="taxon:10090"  
 /clone="A530070C03"  
 /sex="male"  
 /tissue\_type="aorta and vein"  
 /clone\_lib="RIKEN full-length enriched mouse cDNA library"  
 /dev\_stage="adult"  
 1..1292  
 /note="activation-induced cytidine deaminase  
 (MGI:MGI:1342279, GB|NM\_009645, evidence: BLASTN, 99%,  
 match=1048)"

## ORIGIN

Query Match 83.4%; Score 96.8; DB 11; Length 1292;  
 Best Local Similarity 89.7%; Pred. No. 9,1e-21;  
 Matches 104; Conservative 0; Mismatches 12; Indels 0; Gaps 0;  
 QY 1 ATTATTTTACTGCTGAATACCTTTGTAGAAAACACGAAAGACTTCAAGCCTGGG 60  
 Db 520 ACTATTTTACTGCTGAATACCTTTGTAGAAAACGTAAGAACTTCAAGCCTGGG 579  
 QY 61 AAGGCTGCATGAATAATTGCTCTCCAGACAGCTTCGGCGGATCCTTTG 116  
 Db 580 AAGGCTACATGAATAATTGCTCTCCAGACAGCTTCGGCGGATCCTTTG 635

Search completed: March 13, 2004, 00:39:04  
 Job time : 370.498 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 12, 2004, 12:27:59 ; Search time 6838.04 Seconds

(without alignments)  
13767.261 Million cell updates/sec

Title: US-09-966-880A-15

Perfect score: 2172

Sequence: 1 ccccgatagagctgctgata.....agatgacgataaacttg 2172

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:\*

1: gb\_ba:\*

2: gb\_btg:\*

3: gb\_in:\*

4: gb\_cm:\*

5: gb\_ov:\*

6: gb\_pat:\*

7: gb\_ph:\*

8: gb\_pl:\*

9: gb\_pr:\*

10: gb\_ro:\*

11: gb\_sbs:\*

12: gb\_sy:\*

13: gb\_un:\*

14: gb\_vl:\*

15: em\_ba:\*

16: em\_fun:\*

17: em\_hum:\*

18: em\_in:\*

19: em\_mu:\*

20: em\_om:\*

21: em\_or:\*

22: em\_ov:\*

23: em\_pat:\*

24: em\_ph:\*

25: em\_pl:\*

26: em\_ro:\*

27: em\_sbs:\*

28: em\_un:\*

29: em\_vl:\*

30: em\_hes\_hum:\*

31: em\_hes\_in:\*

32: em\_hes\_other:\*

33: em\_hes\_mus:\*

34: em\_hes\_pln:\*

35: em\_hes\_rtd:\*

36: em\_hes\_mam:\*

37: em\_hes\_vrt:\*

38: em\_sy:\*

39: em\_hes\_hum:\*

40: em\_hes\_mus:\*

41: em\_hesg\_other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2172	100.0	2172	6	BD016840 Novel cyt
2	2172	100.0	2791	6	AB040431 Homo sapi
3	2172	100.0	2818	6	BD016833 Novel cyt
4	2172	100.0	6564	6	BD016835 Novel cyt
5	2172	100.0	11204	6	BD016860 Novel cyt
6	2172	100.0	11204	9	AB040430 Homo sapi
7	2172	100.0	71132	9	AC092184 Homo sapi
8	1199.6	55.2	1828	9	BC006296 Homo sapi
9	307	14.1	173071	9	AC009477 Homo sapi
10	305.6	14.1	123101	9	HS108F13 Homo sapi
11	304.6	14.0	135404	9	AC084024 Homo sapi
12	304.6	14.0	174492	2	AC024242 Homo sapi
13	304.6	14.0	195623	2	AP004218 Homo sapi
14	303.6	14.0	172884	2	AC091164 Homo sapi
15	303.6	14.0	176063	2	AC024035 Homo sapi
16	303.6	14.0	194091	9	AP003534 Homo sapi
17	303.6	14.0	195323	2	AP003549 Homo sapi
18	303	14.0	159692	2	AC013693 Homo sapi
19	303	14.0	161806	2	AC016495 Homo sapi
20	303	14.0	183859	9	AC023632 Homo sapi
21	302.8	13.9	131943	9	AC005484 Homo sapi
22	301.4	13.9	183586	9	AC092562 Papio ham
23	301	13.9	234481	2	AC022214 Homo sapi
24	301	13.9	235968	2	AC087393 Homo sapi
25	299.4	13.8	163520	9	AL391259 Human DNA
26	298.8	13.8	183461	9	AC006121 Homo sapi
27	298.2	13.7	180290	9	AC005081 Homo sapi
28	298.2	13.7	204679	9	AC099676 Homo sapi
29	297.6	13.7	111257	9	AL354917 Human DNA
30	297.6	13.7	147885	2	AL353793 Homo sapi
31	297.4	13.7	130855	9	AC004089 Homo sapi
32	297.4	13.7	219935	9	AC005015 Homo sapi
33	297.2	13.7	119794	9	HS295844 Homo sapi
34	297.2	13.7	193608	2	AC019126 Homo sapi
35	296.2	13.6	179372	2	AC040922 Homo sapi
36	295.6	13.6	101722	9	AC132192 Homo sapi
37	295.2	13.6	190076	9	AC008403 Homo sapi
38	294.6	13.6	137506	9	AC004824 Homo sapi
39	294	13.5	143773	9	AC005631 Homo sapi
40	294	13.5	153125	9	AC104456 Homo sapi
41	293.8	13.5	112356	9	AL138812 Human DNA
42	293.8	13.5	132987	9	AC026970 Homo sapi
43	293.8	13.5	153652	2	AC092767 Homo sapi
44	293.8	13.5	159439	9	AC074008 Homo sapi
45	293.8	13.5	169888	2	AC140007 Homo sapi

#### ALIGNMENTS

RESULT 1

BD016840

LOCUS BD016840

DEFINITION Novel cytidine deaminase.

ACCESSION BD016840

VERSION BD016840.1 GI:22558016

KEYWORDS UP 2001245669-A/13.

SOURCE UP 2001245669-A/13.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 2172)

AUTHORS Honjo, T. and Muramatsu, M.

TITLE Novel cytidine deaminase

JOURNAL Patent: JP 2001245669-A 13 11-SEP-2001;



COMMENT  
JAPAN TOBACCO INC. TASUKU HONJO  
OS Homo sapiens (human)  
FN JP 2001245669-A/13  
PD 11-SEP-2001  
PF 28-MAR-2000 JP 2000092981  
PI TASUKU HONJO, MASAMICHI MORIMATSU  
PC C12N15/09, A61K39/395, A61P1/00, A61P11/06, A61P13/12,  
PC A61P17/00,  
PC A61P27/02, A61P27/16, A61P37/02, A61P37/08, C07K16/18, C12N1/19, PC  
C12N1/21,  
PC C12N5/10, C12N9/78, C12P21/02, C12P21/08, C12N1/21, C12R1/19, PC  
C12N5/10, C12R1/91, C12N15/00, C12N5/00, C12N5/00, C12R1/91 CC  
FH Key Location/Qualifiers.  
1. 2172 Location/Qualifiers  
source  
1. 2172  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

ORIGIN

Query Match 100.0%; Score 2172; DB 6; Length 2172;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCCTGATGAGGTGATGACTTACGAGACGATTTGCTGACTTTGGGACTTTGATAGCA 60  
DB 1 CCCCTGATGAGGTGATGACTTACGAGACGATTTGCTGACTTTGGGACTTTGATAGCA 60  
QY 61 CTTCCAGGAATGTCACACAGATGAATACTCTGCTGAGACAGTGTGATAAAAACAGT 120  
DB 61 CTTCCAGGAATGTCACACAGATGAATACTCTGCTGAGACAGTGTGATAAAAACAGT 120  
QY 121 CCTTCAAGTCTTCTCTGTTTATTTATTTCTTCACTCTTCTTGAAGTTTACAGAAAA 180  
DB 121 CCTTCAAGTCTTCTCTGTTTATTTATTTCTTCACTCTTCTTGAAGTTTACAGAAAA 180  
QY 181 ATATTATATACACCTTTTAAAAAGATCTGCTGAAATAGAGAGAACACAGT 240  
DB 181 ATATTATATACACCTTTTAAAAAGATCTGCTGAAATAGAGAGAACACAGT 240  
QY 241 CTGGCAGGAGACGTGCTGCAATTTGTCAGTTTGAATGCAATTTGCCCTACTGGGA 300  
DB 241 CTGGCAGGAGACGTGCTGCAATTTGTCAGTTTGAATGCAATTTGCCCTACTGGGA 300  
QY 301 ATTAACAGAACTGACAGACCTGGAGACCTTAAGTGTCAAGTTTCTATGACTTTTA 360  
DB 301 ATTAACAGAACTGACAGACCTGGAGACCTTAAGTGTCAAGTTTCTATGACTTTTA 360  
QY 361 GGTAGATAGAGACAAAGGTAGATCTTAAAAAGCATGCTGAGAGATCAAAATGTTTTTA 420  
DB 361 GGTAGATAGAGACAAAGGTAGATCTTAAAAAGCATGCTGAGAGATCAAAATGTTTTTA 420  
QY 421 TATCAACATCCTTATTTATTTGATTCATTGATTAAGTAAACAGTGTGATAGATTTT 480  
DB 421 TATCAACATCCTTATTTATTTGATTCATTGATTAAGTAAACAGTGTGATAGATTTT 480  
QY 481 TCTATCTTTTCCCTTGAAGTTTACTTCAAGTAAACAACTCTTCCATCAGGCATGA 540  
DB 481 TCTATCTTTTCCCTTGAAGTTTACTTCAAGTAAACAACTCTTCCATCAGGCATGA 540  
QY 541 TCTATGAGACCTCTTATAGAGATCTGGGTGATTTGACCCCAACATCTTCCAAA 600  
DB 541 TCTATGAGACCTCTTATAGAGATCTGGGTGATTTGACCCCAACATCTTCCAAA 600  
QY 601 GCTTATATATCCATATGAGCTGTATGTTTATATAGAGAGAGATGTTTATGTTTG 660  
DB 601 GCTTATATATCCATATGAGCTGTATGTTTATATAGAGAGAGATGTTTATGTTTG 660  
QY 661 TACAAAGAGATTTGTTATGGGTGGGATGAGGTATAGACATGATGTCACCTTCA 720  
DB 661 TACAAAGAGATTTGTTATGGGTGGGATGAGGTATAGACATGATGTCACCTTCA 720  
QY 721 GCTACTTTTAAATAAGATCTTAAATGGGACGAGAGATGTGAACAAGACCCCTAATPA 780

DB 721 GCTACTTTTAAATAAGATCTTAAATGGGACGAGAGATGTGAACAAGACCCCTAATPA 780  
QY 781 TGGGTGATGTCGAGATGACAAATCTTCTGAGAAAGCAACCTTTTAAAGAAAGCCCT 840  
DB 781 TGGGTGATGTCGAGATGACAAATCTTCTGAGAAAGCAACCTTTTAAAGAAAGCCCT 840  
QY 841 AATTGAGAAACCCCAAACTTCAATATCATATATTTAGCAACAAATGAGAGAAAGTTG 900  
DB 841 AATTGAGAAACCCCAAACTTCAATATCATATATTTAGCAACAAATGAGAGAAAGTTG 900  
QY 901 CTTGATGTTGGGAGAGAGAAATCTATTTGCTCTGCTGGGTCTTCAATCTCAGAAATG 960  
DB 901 CTTGATGTTGGGAGAGAGAAATCTATTTGCTCTGCTGGGTCTTCAATCTCAGAAATG 960  
QY 961 CCAATCAGTCAAGCTTGTCTACATTTTGAATGATGATGATGCTTCTCCCAAGATAT 1020  
DB 961 CCAATCAGTCAAGCTTGTCTACATTTTGAATGATGATGATGCTTCTCCCAAGATAT 1020  
QY 1021 TAACTATATAGAGATGTCACAAAGATGATTAAGCTGCGAACCGTGGACACGCG 1080  
DB 1021 TAACTATATAGAGATGTCACAAAGATGATTAAGCTGCGAACCGTGGACACGCG 1080  
QY 1081 TCATAGTTCTAGCTGTGGAGGTTGAGAGAGAGATGCTTGAACAAGGTGTTCA 1140  
DB 1081 TCATAGTTCTAGCTGTGGAGGTTGAGAGAGAGATGCTTGAACAAGGTGTTCA 1140  
QY 1141 GGGCAGCTGGGCAATTAACAAATGCTCTCTCAAAAAAAAAAAAAAAAAAGAA 1200  
DB 1141 GGGCAGCTGGGCAATTAACAAATGCTCTCTCAAAAAAAAAAAAAAAAAAGAA 1200  
QY 1201 GAGAGAGGCGCGGCGGTGCTCAAGCTGTAATCCAGACTTTGGAGGCGGAGCC 1260  
DB 1201 GAGAGAGGCGCGGCGGTGCTCAAGCTGTAATCCAGACTTTGGAGGCGGAGCC 1260  
QY 1261 GGGGAGATCACCTGTGGTCAAGGATTTGAGACCAAGCTGGCAACATGGCAAAACCCCGT 1320  
DB 1261 GGGGAGATCACCTGTGGTCAAGGATTTGAGACCAAGCTGGCAACATGGCAAAACCCCGT 1320  
QY 1321 CTGACTCAAAATGCAAAATTTAGCCAGGCGGTGAGACGCACTGTAATCCAGCTAC 1380  
DB 1321 CTGACTCAAAATGCAAAATTTAGCCAGGCGGTGAGACGCACTGTAATCCAGCTAC 1380  
QY 1381 TTGGAGAGCTGAGGACAGAGAAATGCTTGAACCCAGAGATGAGGTTTCAATGAGCTGA 1440  
DB 1381 TTGGAGAGCTGAGGACAGAGAAATGCTTGAACCCAGAGATGAGGTTTCAATGAGCTGA 1440  
QY 1441 GATGTCGCTTGAATCCAGCTGGGCGCAAGACCAAGACTGTCTCAGAAAAAAA 1500  
DB 1441 GATGTCGCTTGAATCCAGCTGGGCGCAAGACCAAGACTGTCTCAGAAAAAAA 1500  
QY 1501 AAAAAAAG 1560  
DB 1501 AAAAAAAG 1560  
QY 1561 TGCAAGAAATTTGCTTTATCCACAAATGTAAGAGCAATTAAGGATCCCTATTTTG 1620  
DB 1561 TGCAAGAAATTTGCTTTATCCACAAATGTAAGAGCAATTAAGGATCCCTATTTTG 1620  
QY 1621 TCTCTTTTGTGTCTATTTTGTCCCTTAAACATGCTTGAAGTGAAGAAATATTCAG 1680  
DB 1621 TCTCTTTTGTGTCTATTTTGTCCCTTAAACATGCTTGAAGTGAAGAAATATTCAG 1680  
QY 1681 ATAAACATATCCCTGTGCGCTTATTAAGCAACCTTGAAGTGAAGTGAAGTGAAGTGA 1740  
DB 1681 ATAAACATATCCCTGTGCGCTTATTAAGCAACCTTGAAGTGAAGTGAAGTGAAGTGA 1740  
QY 1741 ACAGAGAACTTGAATGACAACTGTCTTATTTTAACTTTATGATGATGATTTGATA 1800  
DB 1741 ACAGAGAACTTGAATGACAACTGTCTTATTTTAACTTTATGATGATGATTTGATA 1800  
QY 1801 AGATTTAAATTTGTTACTTCAATGATGATTTATATTTTATTTTGGGTCTAAG 1860



FEATURES	source
gene	AB040431
CDS	AB040431
DEFINITION	AB040431 Homo sapiens AID mRNA for activation-induced cytidine deaminase, complete CDS.
ACCESSION	AB040431
VERSION	AB040431.1 GI:9988409
KEYWORDS	AID; activation-induced cytidine deaminase; Human AID.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (sites)
TITLE	Muto, T., Muramatsu, M., Tanikawa, M., Kinoshita, K. and Honjo, T.
JOURNAL	Isolation, tissue distribution, and chromosomal localization of the
MEDLINE	human activation-induced cytidine deaminase (AID) gene
PUBMED	Genomics 68 (1), 85-88 (2000)
REFERENCE	20408890
AUTHORS	10950930
REFERENCE	2 (sites)
AUTHORS	Remy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sarai, O.,
TITLE	Catalan, N., Forveille, M., Dufourcq-Lageouse, R., Gennery, A.,
JOURNAL	Tecsan, I., Essoy, F., Kayseril, H., Ugazio, A.G., Bronsse, N.,
MEDLINE	Mutamat, S., Notarangelo, L.D., Kinoshita, K., Honjo, T., Fischer, A.
PUBMED	and Durandy, A.
REFERENCE	Activation-induced cytidine deaminase (AID) deficiency causes the
AUTHORS	autosomal recessive form of the Hyper-IgM syndrome (HIGM2)
TITLE	Ce11 102 (5), 565-575 (2000)
JOURNAL	20460541
MEDLINE	11007475
PUBMED	3 (bases 1 to 2791)
REFERENCE	Muto, T., Muramatsu, M., Tanikawa, M., Kinoshita, K. and Honjo, T.
AUTHORS	Direct Submission
TITLE	Submitted (18-MAR-2000) Tasuku Honjo, Kyoto University, Department
JOURNAL	of Medical Chemistry, Faculty of Medicine, Yoshida, Sakyo-ku,
MEDLINE	Kyoto, Kyoto 606-8501, Japan (E-mail: honjo@omfou.med.kyoto-u.ac.jp,
PUBMED	Tel: 81-75-753-4371 (ex 4371), Fax: 81-75-753-4388)
REFERENCE	Location/Qualifiers
AUTHORS	1..2791
TITLE	/organism="Homo sapiens"
JOURNAL	/mol_type="mRNA"
MEDLINE	/db_xref="taxon:9606"
PUBMED	1.._2791
REFERENCE	/gene="AID"
AUTHORS	77.._673
TITLE	/gene="AID"

		/codon_start=1	
		/product="activation-induced cytidine deaminase"	
		/protein_id="BAB12721.1"	
		/db_xref="GI:9988410"	
		/translation="MDSLIMRRKFLYQFNVMWAKGRRETYLYVKKRDSATSESLDFGLRNKGCHVELFLRYISDMIDLPGRCYRLWTSPSCDCAHADFLKGNP NLSIRFTALYLECEDRKAPDEGRRLRHAGVOAIIPTFSDFYCWNTFVENHERTEPK AMEGLHENSVRLSQLRILLPLYEVDDLDAFRITGLT"	
ORIGIN			
Query Match	100.0%; Score 2172; DB 9;	Length 2791;	
Best Local Similarity	100.0%; Pred. No. 0;		
Matches 2172; Conservative	0; Mismatches	0; Indels	0; Gaps 0;
QY	1	CCCTGTATGAGGTTGATGACTTGACGAGCGCATTCGTAATTGGACCTTGATGCAA	60
DB	620	CCCCGTATGAGGTTGATGACTTGACGAGCGCATTCGTAATTGGACCTTGATGCAA	679
QY	61	CTTCACAGGAATGTACAACAGATAAATATCTCGCTGGAAGACAGTGGATAAAACAGT	120
DB	680	CTTCACAGGAATGTACAACAGATAAATATCTCGCTGGAAGACAGTGGATAAAACAGT	739
QY	121	CCTTCAAGTCTTCTGTTTTTATCTTCAAACCTCTCACTTCTTAGAGTTACAGAAAA	180
DB	740	CCTTCAAGTCTTCTGTTTTTATCTTCAAACCTCTCACTTCTTAGAGTTACAGAAAA	799
QY	181	AATATTATATACACTCTTTAAAAAAGATCTATGCTTGAAATATAGAGAAGAACACAGT	240
DB	800	AATATTATATACACTCTTTAAAAAAGATCTATGCTTGAAATATAGAGAAGAACACAGT	859
QY	241	CTGGCCAGGACGTCGTGCCAATTGGTGCAGTTTGAATGCAACATTGTCCCCTACTGGGA	300
DB	860	CTGGCCAGGACGTCGTGCCAATTGGTGCAGTTTGAATGCAACATTGTCCCCTACTGGGA	919
QY	301	ATAACAGAACTGACGACCTGGGAGCATCCTAAAGSTCACAGTTTTCTATGACTTTTA	360
DB	920	ATAACAGAACTGACGACCTGGGAGCATCCTAAAGSTCACAGTTTTCTATGACTTTTA	979
QY	361	GGINAGATGAGACGAGAAGGTAGATCCTAAAAAAGCATGGTGAGAGATCAAAATGTTTTA	420
DB	980	GGINAGATGAGACGAGAAGGTAGATCCTAAAAAAGCATGGTGAGAGATCAAAATGTTTTA	1039
QY	421	TATGAAATCTCTTATTTATTTGATTCATTTGAGTTAACAGTGGTGTATGATATGATTTT	480
DB	1040	TATGAAATCTCTTATTTATTTGATTCATTTGAGTTAACAGTGGTGTATGATATGATTTT	1099
QY	481	TCTATTCCTTTCCCTTGACGTTTACTTCAAGTAAACAAACTCTTCATCAGGCCATGA	540
DB	1100	TCTATTCCTTTCCCTTGACGTTTACTTCAAGTAAACAAACTCTTCATCAGGCCATGA	1159
QY	541	TCTATAGACCTCCTTAATGAGAGATATCGGTGTGTTGTGACCCCAACCATCTCTCCAAA	600
DB	1160	TCTATAGACCTCCTTAATGAGAGATATCGGTGTGTTGTGACCCCAACCATCTCTCCAAA	1219
QY	601	GCATTAAATATCAATCATGCGCTGTATGTTTATATCAGCAAGACATGTTTTATGTTTG	660
DB	1220	GCATTAAATATCAATCATGCGCTGTATGTTTATATCAGCAAGACATGTTTTATGTTTG	1279
QY	661	TACAAAGAAGATTTGTAATGGGTGGGATGGAAGTATAGACATGATGATGATCAGCTTCAA	720
DB	1280	TACAAAGAAGATTTGTAATGGGTGGGATGGAAGTATAGACATGATGATGATCAGCTTCAA	1339
QY	721	GCTACTTAAATTAAGATCTTAAATATGGGACGAGACTGTGAACAGACACCTTAATAA	780
DB	1340	GCTACTTAAATTAAGATCTTAAATATGGGACGAGACTGTGAACAGACACCTTAATAA	1399
QY	781	TGGGTGATGTCTGAAGTAGCAAAATCTTCTGAAAAGCAAACTCTTTTATAGAAATGTCCT	840
DB	1400	TGGGTGATGTCTGAAGTAGCAAAATCTTCTGAAAAGCAAACTCTTTTATAGAAATGTCCT	1459
QY	841	AATTTAAGAAACCCCAACACTTCACATATCAATATTAAGAAACAATTTGAAAGAAAGTGG	900
DB	1460	AATTTAAGAAACCCCAACACTTCACATATCAATATTAAGAAACAATTTGAAAGAAAGTGG	1519



QY	901	CTTGATGTTGGGAGAGGAAATCATATGGCTCGTGGGCTCTCTCATCTCAAAATG	960
Db	1520	CTTGATGTTGGGAGAGGAAATCATATGGCTCTCTGGCTCTCTCATCTCAAAATG	1578
QY	961	CCAAATCAGGTCAGGTTTGCTACATTTGTATGTGTGATGCTTCTCCAAAGGTATAT	1020
Db	1580	CCAAATCAGGTCAGGTTTGCTACATTTGTATGTGTGTATGCTTCTCCAAAGGTATAT	1633
QY	1021	TAACTATATTAAGAAGTTGTGACAAAACGAATGATTAAGCTGCGAACCGTGGCACAGC	1086
Db	1640	TAACTATATTAAGAAGTTGTGACAAAACGAATGATTAAGCTGCGAACCGTGGCACAGC	1699
QY	1081	TCATGTTCTTACTCTCTGGGAGGTTGAGAGGAGGAGATGGCTTGACACAGTGTTCAA	1144
Db	1700	TCATGTTCTTACTCTCTGGGAGGTTGAGAGGAGGAGATGGCTTGACACAGTGTTCAA	1755
QY	1141	GCGCAGCCTGGGCAACATPACAGATCCGTCTCTCAAAAAAAAAAAAAAAAAAGAA	1200
Db	1760	GCGCAGCCTGGGCAACATPACAGATCCGTCTCTCAAAAAAAAAAAAAAAAAAGAA	1815
QY	1201	GAGAGAGGCGCGGCGGTGTGCTCACGCTGTAAATCCAGACATTTGGAGCGCGAGCC	1266
Db	1820	GAGAGAGGCGCGGCGGTGTGCTCACGCTGTAAATCCAGACATTTGGAGCGCGAGCC	1879
QY	1261	GGGCGGATCATCTGTGTGTGTCAGAGTTTGACACAGCTGTGGCAATGGCAAAACCCGT	1322
Db	1880	GGGCGGATCATCTGTGTGTGTCAGAGTTTGACACAGCTGTGGCAATGGCAAAACCCGT	1933
QY	1321	CTGTACTCAAAATGCAAAATTAAGCGAGCGGTGTAGACAGCACTGTAAATCCAGCTAC	1386
Db	1940	CTGTACTCAAAATGCAAAATTAAGCGAGCGGTGTAGACAGCACTGTAAATCCAGCTAC	1999
QY	1381	TTGGAGGCTGAGCAGAGAAATCCGCTTGAAACCCAGAGGTGAGGTTGCAATAAGCTGA	1440
Db	2000	TTGGAGGCTGAGCAGAGAAATCCGCTTGAAACCCAGAGGTGAGGTTGCAATAAGCTGA	2055
QY	1441	GATCGTCCGCTTGCACTCCAGCTGTGGCGACAAAGCAAGACTGTCTCAGAAAAAAAA	1500
Db	2060	GATCGTCCGCTTGCACTCCAGCTGTGGCGACAAAGCAAGACTGTCTCAGAAAAAAAA	2119
QY	1501	AAAAAAAAAGAGAGAGAGAGAAABAGAACATTTTGGAGAGAGATGGGAGAGCAT	1566
Db	2120	AAAAAAAAAGAGAGAGAGAGAAABAGAACATTTTGGAGAGAGATGGGAGAGCAT	2179
QY	1561	TGCAAGAGAAATGTGCTTTATCCACAACAAATTTAAGAGCCATAAGGATCCCATTTG	1622
Db	2180	TGCAAGAGAAATGTGCTTTATCCACAACAAATTTAAGAGCCATAAGGATCCCATTTG	2239
QY	1621	TCTCTTTGGTGTCTATTTGTCCCTPACACATGCTTGTGACAGTAGAGAAATATTCGA	1680
Db	2240	TCTCTTTGGTGTCTATTTGTCCCTPACACATGCTTGTGACAGTAGAGAAATATTCGA	2299
QY	1661	ATAACCATATCCCTGTGCGCTTATTACTAGAACCCCTTGCAATGAAATGACAGATCC	1744
Db	2300	ATAACCATATCCCTGTGCGCTTATTACTAGAACCCCTTGCAATGAAATGACAGATCC	2359
QY	1741	ACAGAGAAACTGGATGACAACTGTCTTAATTTTAATCTTAATGACATAAGTTGTAA	1800
Db	2360	ACAGAGAAACTGGATGACAACTGTCTTAATTTTAATCTTAATGACATAAGTTGTAA	2419
QY	1801	AGAGTTAAAAATGTACTTCATGATTCATTTATATTTAATAATTTGGCTATAG	1860
Db	2420	AGAGTTAAAAATGTACTTCATGATTCATTTATATTTAATATTTTGGCTATAG	2479
QY	1861	ATTTTATTAACATGATTTCTTTTCTGATATATGAATGAGATCCAAAGCTTCA	1920
Db	2480	ATTTTATTAACATGATTTCTTTTCTGATATATGAATGAGATCCAAAGCTTCA	2539
QY	1921	AATTTATTAACATTTAGAAATGATCTTAATPACACAGTATGTAATTCAGTAT	1986
Db	2540	AATTTATTAACATTTAGAAATGATCTTAATPACACAGTATGTAATTCAGTAT	2599

QY		1981	GATGGTACGAAGCCATTCTCTGTGATTTTGTAACCTTTATAGAAGCAATTTGCTC	2046
Db		2600	GGTGCTAGGAAGCCATTCTCTGTGATTTTGTAACCTTTATAGAAGCAATTTGCTTC	2658
QY		2041	TGGCTCACTTTCATCACTTAAATAAAGATAAATAATTTTGAAGCTGTGAAGATAAA	2100
Db		2660	TGGCTCACTTTCATCACTTAAATAAAGATAAATAATTTTGAAGCTGTGAAGATAAA	2719
QY		2101	TACCAATATAAATATATATAAAAGCATTTATATGAAGTTAAAAATCAGTAGAT	2160
Db		2720	TACCAATATAAATATATATAAAAGCATTTATATGAAGTTAAAAATCAGTAGAT	2779
QY		2161	GAATAAACCTTG 2172	
Db		2780	GAATAAACCTTG 2791	
RESULT 3				
BD016833				
LOCUS	BD016833	2818 bp	DNA	linear PAT 27-AUG-2002
DEFINITION	Novel cytidine deaminase.			
ACCESSION	BD016833			
VERSION	BD016833.1 GI:22558009			
KEYWORDS	JP 2001245669-A/6.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE	Novel cytidine deaminase			
AUTHORS	Patent: JP 2001245669-A 6 11-SEP-2001;			
JOURNAL	JAPAN TOBACCO INC,TASUJU HONOJO			
COMMENT	OS Homo sapiens (human) PN JP 2001245669-A/6 PD 11-SEP-2001 PF 28-MAR-2000 JP 2000052981 PI TASUJU HONOJO,MASAMICHI MURAMATSU PC C12N15/09,A6IK39/395,A6IK39/395,A6IP1/00,A6IP12/06,A6IP13/12, PC A6IP17/00, PC A6IP27/02,A6IP27/16,A6IP37/02,A6IP37/08,C07K16/18,C12N1/19, PC C12N1/21 PC C12N5/10,C12N5/78,C12P21/02,C12P21/08/(C12N1/21,C12R1:19), PC (C12N5/10,C12R1:91),C12N15/00,C12N5/00,(C12N5/00,C12R1:91) CC FH Key Location/Qualifiers FT 5'UTR (1)..(179) FT CDS (80)..(676) FT 3'UTR (677)..(2818).			
FEATURES	Location/Qualifiers			
source	1..2818 /organism="Homo sapiens" /mol_type="genomic DNA" /db_xref="taxon:9606"			
ORIGIN				
Query Match	100.0%; Score 2172; DB 6; Length 2818;			
Best Local Similarity	100.0%; Pred. No. 0;			
Matches 2172; Conservative	0; Mismatches 0; Indels 0; Gaps 0;			
QY		1	CCCCGTGATGAGTTGATGACTTACGAGACGCCATTTGTACTTTGGGACCTTGATAGCA	60
Db		623	CCCCGTGATGAGTTGATGACTTACGAGACGCCATTTGTACTTTGGGACCTTGATAGCAA	682
QY		61	CTTCCAGGAATGCACAACAGATGAATATCTCGCTGGAAGACAGTGGATPAAAAAAGT	120
Db		683	CTTCCAGGAATGTACACAACGATTAATATCTCGCTGGAAGACAGTGGATPAAAAAAGT	742
QY		121	CCTTCAAGTCTTCTGTTTTATCTTCAACTCTCACTTCTTGAAGATTACAGAAAA	180
Db		743	CCTTCAAGTCTTCTGTTTTATCTTCAACTCTCACTTCTTGAAGATTACAGAAAA	802
QY		181	ATATTATATACGACTCTTTAAAAAGATCTATGCTTGAATAATAGAGAAACACAGGT	240



Db 803 ATATTATATAGAGCTCTTTAAAAAGATCTATGCTTGAATAATAGAGAGGACACAGGT 862  
 Qy 241 CTGGCCAGGAGCGTGTCCAAATTGGTGAAGTTTGAATGCAACATGGCCCTTACCTGGGA 300  
 Db 863 CTGGCCAGGAGCGTGTCCAAATTGGTGAAGTTTGAATGCAACATGGCCCTTACCTGGGA 922  
 Qy 301 ATAAAGAACTGACAGGAGCTGGGAGCATCTTAAAGTGCACAGTTTTTCTATGACTTTTA 360  
 Db 923 ATAAAGAACTGACAGGAGCTGGGAGCATCTTAAAGTGCACAGTTTTTCTATGACTTTTA 982  
 Qy 361 GGTAGAGTGAAGAGAGAGTATCTTAAAAAGCATGGTGAAGAGATCAATATGTTTTTA 420  
 Db 983 GGTAGAGTGAAGAGAGAGTATCTTAAAAAGCATGGTGAAGAGATCAATATGTTTTTA 1042  
 Qy 421 TATCAACATCCTTTATTTATTTGATTCATTTGAGTTTAAAGTGGTTATGATGATTTT 480  
 Db 1043 TATCAACATCCTTTATTTATTTGATTCATTTGAGTTTAAAGTGGTTATGATGATTTT 1102  
 Qy 481 TCTATCTTTTCCCTTGAAGTTTACTTTCAGTAAACAAACTCTTCATCAGGCGATGA 540  
 Db 1103 TCTATCTTTTCCCTTGAAGTTTACTTTCAGTAAACAAACTCTTCATCAGGCGATGA 1162  
 Qy 541 TCTATAGGAGCTCTTATAGAGATCTGGGTGATTTGACCCCAACCATCTCTCCAA 600  
 Db 1163 TCTATAGGAGCTCTTATAGAGATCTGGGTGATTTGACCCCAACCATCTCTCCAA 1222  
 Qy 601 GCATTATATCCCAATCATGCGCTGTATGTTTATATCAGCAAGAGATGTTTTATGTTG 660  
 Db 1223 GCATTATATCCCAATCATGCGCTGTATGTTTATATCAGCAAGAGATGTTTTATGTTG 1282  
 Qy 661 TCAAAAGAAAGTTGTTATGAGTGGGAGTGAAGATGATGACCAAGTGCACCTTCAA 720  
 Db 1283 TCAAAAGAAAGTTGTTATGAGTGGGAGTGAAGATGATGACCAAGTGCACCTTCAA 1342  
 Qy 721 GCTACTTTATTAAGGATCTTAAAAAGGAGAGAGAGATGATGACCAAGTGCACCTTCAA 780  
 Db 1343 GCTACTTTATTAAGGATCTTAAAAAGGAGAGAGATGATGACCAAGTGCACCTTCAA 1402  
 Qy 781 TGGGTGATGCTGTAAGTGAAGTCTTGTGAAAAGCAACTCTTTTAAAGAGTCTCCT 840  
 Db 1403 TGGGTGATGCTGTAAGTGAAGTCTTGTGAAAAGCAACTCTTTTAAAGAGTCTCCT 1462  
 Qy 841 AATTAGAAACACCAACAATCTTCAATATCAATATGCAAAACATTTGAAAGAGTTG 900  
 Db 1463 AATTAGAAACACCAACAATCTTCAATATGCAAAACATTTGAAAGAGTTG 1522  
 Qy 901 CTTGAATGTTGGGAGAGAGAAATCTATTTGCTCTGAGGCTCTTCACTCAAAATG 960  
 Db 1523 CTTGAATGTTGGGAGAGAGAAATCTATTTGCTCTGAGGCTCTTCACTCAAAATG 1582  
 Qy 961 CCAATCAGGTCAAGTTGCTCAATTTGTATGTTGATGCTCTTCCCAAGGTATAT 1020  
 Db 1583 CCAATCAGGTCAAGTTGCTCAATTTGTATGTTGATGCTCTTCCCAAGGTATAT 1642  
 Qy 1021 TAACTATATAGAGAGTGTGACAAAACAAATGATTAAGCTGCGAACCTGGGACACGC 1080  
 Db 1643 TAACTATATAGAGAGTGTGACAAAACAAATGATTAAGCTGCGAACCTGGGACACGC 1702  
 Qy 1081 TCATAGTTCTAGTGTGTTGGAGGTTGAGAGAGGAGAGATGGCTTGAAACAGAGTGTCAA 1140  
 Db 1703 TCATAGTTCTAGTGTGTTGGAGGTTGAGAGAGGAGAGATGGCTTGAAACAGAGTGTCAA 1762  
 Qy 1141 GGCACACCTGGGACAAATTAACAAGATCTGTCTTCAAAAAAAGAAAAAAGAAA 1200  
 Db 1763 GGCACACCTGGGACAAATTAACAAGATCTGTCTTCAAAAAAAGAAAAAAGAAA 1822  
 Qy 1201 GAGAGAGGCGCGGCGTGTGCTCAGCCTGTATCCAGAGACTTTGGGAGCGGAGGC 1260  
 Db 1823 GAGAGAGGCGCGGCGTGTGCTCAGCCTGTATCCAGAGACTTTGGGAGCGGAGGC 1882  
 Qy 1261 GGGCGGATCACTGTGGTCAAGAGTTTGAGACCAAGCTGGCCAAACATGGCAAAACCCGCT 1320  
 Db 1883 GGGCGGATCACTGTGGTCAAGAGTTTGAGACCAAGCTGGCCAAACATGGCAAAACCCGCT 1942

Qy 1321 CTGTACTCAAAAATGCAAAAATTAGCCAGGCGTGTAGAGGAGCACTGTATCCAGCTAC 1380  
 Db 1943 CTGTACTCAAAAATGCAAAAATTAGCCAGGCGTGTAGAGGAGCACTGTATCCAGCTAC 2002  
 Qy 1381 TTGGAGGCTGAGGAGAGAGAAATCGCTTGAATCCAGAGAGTGAAGTTGCAATGCTGA 1440  
 Db 2003 TTGGAGGCTGAGGAGAGAGAAATCGCTTGAATCCAGAGAGTGAAGTTGCAATGCTGA 2062  
 Qy 1441 GATCGTCGGCTTGCACTCCAGGCTGGGAGCAAGAGCAAGATCTGTCTCAGAAAAAAA 1500  
 Db 2063 GATCGTCGGCTTGCACTCCAGGCTGGGAGCAAGAGCAAGATCTGTCTCAGAAAAAAA 2122  
 Qy 1501 AAAAAAAG 1560  
 Db 2123 AAAAAAAG 2182  
 Qy 1561 TGCAGAGAAATTTGCTTTATTCACAAAATGTAAGAGCCAAATTAAGGATCCCTATTTG 1620  
 Db 2183 TGCAGAGAAATTTGCTTTATTCACAAAATGTAAGAGCCAAATTAAGGATCCCTATTTG 2242  
 Qy 1621 TCTCTTTGGTGTCTATTTGTCCCTTAAACAATGCTTTGACAGTGAAGAAATATTTGCA 1680  
 Db 2243 TCTCTTTGGTGTCTATTTGTCCCTTAAACAATGCTTTGACAGTGAAGAAATATTTGCA 2302  
 Qy 1681 ATAAACATATCCCTGTCGGCTTATTACCTAGAGACCCCTGCAATGAGATGACAGATCC 1740  
 Db 2303 ATAAACATATCCCTGTCGGCTTATTACCTAGAGACCCCTGCAATGAGATGACAGATCC 2362  
 Qy 1741 ACAGAGAAATCTGAAGACCAACTGTCTATTTTAACTTATTTGAACATTAAGTTGTAAA 1800  
 Db 2363 ACAGAGAAATCTGAAGACCAACTGTCTATTTTAACTTATTTGAACATTAAGTTGTAAA 2422  
 Qy 1801 AGAGTAAAAATTTGTTACTTATCATGATATCATTTATTTATTTATTTTGGCTTAATG 1860  
 Db 2423 AGAGTAAAAATTTGTTACTTATCATGATATCATTTATTTATTTATTTTGGCTTAATG 2482  
 Qy 1861 ATTTTATTTATTAATGATTTTCTTTCTGTATATTTGAATGAGAGCTCAAGCTTCATA 1920  
 Db 2483 ATTTTATTTATTAATGATTTTCTTTCTGTATATTTGAATGAGAGCTCAAGCTTCATA 2542  
 Qy 1921 AATTATTAATTTAGAAATGATCTATTAACAACATGATTAATTTGTAATGACAGTAAT 1980  
 Db 2543 AATTATTAATTTAGAAATGATCTATTAACAACATGATTAATTTGTAATGACAGTAAT 2602  
 Qy 1981 GGTGCTACAGAGCAATTTCTGTTGATTTTATGTAACCTTTATGACAGCAAAATTTGCTTC 2040  
 Db 2603 GGTGCTACAGAGCAATTTCTGTTGATTTTATGTAACCTTTATGACAGCAAAATTTGCTTC 2662  
 Qy 2041 TGGCTCACTTCAATCAATGATTAATTAATGATTAATTTTGGAGCTGTGAAGATTA 2100  
 Db 2663 TGGCTCACTTCAATCAATGATTAATTAATGATTAATTTTGGAGCTGTGAAGATTA 2722  
 Qy 2101 TACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2160  
 Db 2723 TACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2782  
 Qy 2161 GGAATTAACCTTG 2172  
 Db 2783 GGAATTAACCTTG 2794

RESULT 4  
 BD016835  
 LOCUS BD016835  
 DEFINITION Novel cytidine deaminase.  
 ACCESSION BD016835.1 GI:22558011  
 VERSION JP 2001245669-A/8.  
 KEYWORDS Homo sapiens (human)  
 SOURCE  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.



```

REFERENCE 1 (bases 1 to 6564)
AUTHORS Honjo, T. and Muramatsu, M.
TITLE Novel cyclidine deaminase
JOURNAL Patent: JP 2001245669-A 8 11-SEP-2001;
        JAPAN TOBACCO INC, TASUKU HONJO
COMMENT OS Homo sapiens (human)
        PN JP 2001245669-A/8
        PD 11-SEP-2001
        PE 28-MAR-2000 JP 2000092981
        PI TASUKU HONJO, MASAMICHI MURAMATSU
        PC C12N5/09, A61K39/395, A61K39/395, A61P1/00, A61P11/06, A61P13/12,
        PC A61P17/00,
        PC A61P27/02, A61P27/16, A61P37/02, A61P37/08, C07K16/18, C12N1/19, PC
        C12N1/21,
        PC C12N5/10, C12N9/78, C12P21/02, C12P21/08// (C12N1/21, C12R1:19), PC
        (C12N5/10, C12R1:91), C12N5/00, C12N5/00, (C12N5/00, C12R1:91) CC
FEATURES
source          Location/Qualifiers
                 1..6564
                 /organism="Homo sapiens"
                 /mol_type="genomic DNA"
                 /db_xref="taxon:9606"
ORIGIN
Query Match      100.0%; Score 2172; DB 6; Length 6564;
Beet Local Similarity 100.0%; Pred. No. 0;
Matches 2172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1  CCCCTGATGAGGTGATGACTTACGAGACGATTCCTGACTTTGGGACTTTGATAGCAA 60
Db  3741 CCCCTGATGAGGTGATGACTTACGAGACGATTCCTGACTTTGGGACTTTGATAGCAA 3800
QY  61  CTTCCAGGATGTCAACAAGATATATCTCTCTGAGACAGACAGTATAAAAAGT 120
Db  3801 CTTCCAGGATGTCAACAAGATATATCTCTCTGAGACAGACAGTATAAAAAGT 3860
QY  121 CCTTAAGCTCTCTGTTTATCTCAACTCACTTCTTGAAGTTTACAGAAAA 180
Db  3861 CCTTAAGCTCTCTGTTTATCTCAACTCACTTCTTGAAGTTTACAGAAAA 3920
QY  181 ATATTATATAGACTCTTTAAAAAGATCTATGCTTGAAAAATAGAGAGACAAGT 240
Db  3921 ATATTATATAGACTCTTTAAAAAGATCTATGCTTGAAAAATAGAGAGACAAGT 3980
QY  241 CTGGCAGGAGCGTGTGCAATTGTGCAAGTTTGAATGCAACATTTGCCCTACTGGGA 300
Db  3981 CTGGCAGGAGCGTGTGCAATTGTGCAAGTTTGAATGCAACATTTGCCCTACTGGGA 4040
QY  301 ATTAACAGAACTGCAAGACCTGGGAGACATCTAAAGTCCAAAGCTTTTCTATGACTTTA 360
Db  4041 ATTAACAGAACTGCAAGACCTGGGAGACATCTAAAGTCCAAAGCTTTTCTATGACTTTA 4100
QY  361 GGTAGAGTGAAGAGAGAGTATGCTTAAAAAGCATGGTGAAGAGATCAATGTTTTTA 420
Db  4101 GGTAGAGTGAAGAGAGAGTATGCTTAAAAAGCATGGTGAAGAGATCAATGTTTTTA 4160
QY  421 TATCAACATCCTTATATTATTAATCAATTTAGTTAACAAGTGTATAGATAGATTTT 480
Db  4161 TATCAACATCCTTATATTATTAATCAATTTAGTTAACAAGTGTATAGATAGATTTT 4220
QY  481 TCTATTCTTTTCCCTTGAAGTTTACTTTCAAGTAACAACAACCTTCCATCAGGCGAAGA 540
Db  4221 TCTATTCTTTTCCCTTGAAGTTTACTTTCAAGTAACAACAACCTTCCATCAGGCGAAGA 4280
QY  541 TCTATGAGACCTCTAATGAGATCTGGGTGATGTGACCCCAACCATCTCCCAA 600
Db  4281 TCTATGAGACCTCTAATGAGATCTGGGTGATGTGACCCCAACCATCTCTCCCAA 4340
QY  601 GCATTATATCCAAATCATGCGCTGTATGTTTAAATCAGACAGAGATGTTTATGTTTG 660
Db  4341 GCATTATATCCAAATCATGCGCTGTATGTTTAAATCAGACAGAGATGTTTATGTTTG 4400
QY  661 TACAAAAGAAATGTTATGGGTGGGAGTGAAGCTATAGACCATGATGCTCACTTCAA 720

```

```

Db  4401 TACAAAAGAAATGTTATGGGTGGGAGTGAAGCTATAGACCATGATGCTCACTTCAA 4460
QY  721 GCTACTTTATATAGATCTTTAAATGGGACAGAGACTGTGAACAAACACCCCTATATA 780
Db  4461 GCTACTTTATATAGATCTTTAAATGGGACAGAGACTGTGAACAAACACCCCTATATA 4520
QY  781 TGGGTGATGTCTGAAGTGAACAATCTTCTGAAAACGCAAACTCTTTTAAAGAACTCCT 840
Db  4521 TGGGTGATGTCTGAAGTGAACAATCTTCTGAAAACGCAAACTCTTTTAAAGAACTCCT 4580
QY  841 AATTGAAACCCCAACAACTTCAATCTATTTAGCAAAATTTGAAAGAAAGTTG 900
Db  4581 AATTGAAACCCCAACAACTTCAATCTATTTAGCAAAATTTGAAAGAAAGTTG 4640
QY  901 CTTGAATGTTGGGAGAGAGAAATCTATTTGGCTCGTGGGCTCTTCATCTCAGAAATG 960
Db  4641 CTTGAATGTTGGGAGAGAGAAATCTATTTGGCTCGTGGGCTCTTCATCTCAGAAATG 4700
QY  961 CCAATCAGGTCAAGGTTTGTCTCAATTTGTATGTGTGATGCTTCTCCAAAGGTATAT 1020
Db  4701 CCAATCAGGTCAAGGTTTGTCTCAATTTGTATGTGTGATGCTTCTCCAAAGGTATAT 4760
QY  1021 TAACTATATAGAGAGTGTGCAAAAAGATTAAGCTGCAACCGTGGCAACGC 1080
Db  4761 TAACTATATAGAGAGTGTGCAAAAAGATTAAGCTGCAACCGTGGCAACGC 4820
QY  1081 TCAATGTTTACCTGCTTGGAGAGTTGAGAGAGAGAGATGCTTGAACAACGTTGCAA 1140
Db  4821 TCAATGTTTACCTGCTTGGAGAGTTGAGAGAGAGAGATGCTTGAACAACGTTGCAA 4880
QY  1141 GGCCAGCTGGGCAATATAAAGATCTGTCTCAAAAAAAAAAAAAAAAAAAAAA 1200
Db  4881 GGCCAGCTGGGCAATATAAAGATCTGTCTCTCAAAAAAAAAAAAAAAAAAAAAA 4940
QY  1201 GAGAGAGGCGGGCGTGTGCTCAAGCTGTAAATCCAGACATTTGGAGGCGGAGCC 1260
Db  4941 GAGAGAGGCGGGCGTGTGCTCAAGCTGTAAATCCAGACATTTGGAGGCGGAGCC 5000
QY  1261 GGGCGGATCACTGTGTGTCAAGAGTTTGAACAACGCTGGCCAAATGGCAAAACCCGT 1320
Db  5001 GGGCGGATCACTGTGTGTCAAGAGTTTGAACAACGCTGGCCAAATGGCAAAACCCGT 5060
QY  1321 CTGTACTCAAAATGCAAAATTTAGACGAGCGTGTAGACGAGCACTGTATCCAGACTAC 1380
Db  5061 CTGTACTCAAAATGCAAAATTTAGACGAGCGTGTAGACGAGCACTGTATCCAGACTAC 5120
QY  1381 TTGGAGGCTGAGGACAGAGAAATCGCTTGAACCCAGAGAGTGAAGTTGCAATAGCTGA 1440
Db  5121 TTGGAGGCTGAGGACAGAGAAATCGCTTGAACCCAGAGAGTGAAGTTGCAATAGCTGA 5180
QY  1441 GATGTGCGCTGTGCACTCAAGCTGGGCGCAAGAGCAACATCTGTCTCAAAAAAAAA 1500
Db  5181 GATGTGCGCTGTGCACTCAAGCTGGGCGCAAGAGCAACATCTGTCTCAAAAAAAAA 5240
QY  1501 AAAAAAAAAAGAGAGAGAGAGAAAGAAACAATTTTGGAGAGAGAGATGGGAGAGCAT 1560
Db  5241 AAAAAAAAAAGAGAGAGAGAGAAAGAAACAATTTTGGAGAGAGAGATGGGAGAGCAT 5300
QY  1561 TGCAAGAAATTTGCTTTATCCAAACAAATGTGAAGACCAATAGAGATCCCTATTTTG 1620
Db  5301 TGCAAGAAATTTGCTTTATCCAAACAAATGTGAAGACCAATAGAGATCCCTATTTTG 5360
QY  1621 TCTCTTTTGTGTCTATTTGTCCCTAACAACCTGCTTTGAACGTGAAGAAATATTCAGA 1680
Db  5361 TCTCTTTTGTGTCTATTTGTCCCTAACAACCTGCTTTGAACGTGAAGAAATATTCAGA 5420
QY  1681 ATTAACATATCCCTGTGCGCTTATTAACCTAGCAACCTTGCATATGAAGATGACAGATCC 1740
Db  5421 ATTAACATATCCCTGTGCGCTTATTAACCTAGCAACCTTGCATATGAAGATGACAGATCC 5480
QY  1741 ACAGAAACCTTGAATGCAACATGTCTTATTTAATCTTATGTCATAGTTGTAAA 1800

```



Db 5481 ACAGGAACTGATGACCAACTGCTGATTTTATCTTATGTATGACATTAAGTTGTAAA 5540  
Qy 1801 AGAGTAAATAATGTTACTTCATGATGATTCATTTATATTTATTTATTTTGGCTAAAG 1860  
Db 5541 AAGAGTAAATAATGTTACTTCATGATGATTCATTTATATTTATTTATTTTGGCTAAAG 5600  
Qy 1861 ATTTTATTTATACATGATTTCTTTTCTGATATATTGAAATGAGCTCGAAGCTTCATA 1920  
Db 5601 ATTTTATTTATACATGATTTCTTTTCTGATATATTGAAATGAGCTCGAAGCTTCATA 5660  
Qy 1921 AATTATTAAGTTAGAAATGATTTCTATATACAAAGTATGATTTGTAACATTTGACAT 1980  
Db 5661 AATTATTAAGTTAGAAATGATTTCTATATACAAAGTATGATTTGTAACATTTGACAT 5720  
Qy 1981 GGTGTACAGAGCAATTTCTGATTTTGTAGTAACTTTTATGACAGAAATTTGCTTC 2040  
Db 5721 GGTGTACAGAGCAATTTCTGATTTTGTAGTAACTTTTATGACAGAAATTTGCTTC 5780  
Qy 2041 TGGCTCACTTCATGATGATTAATAATGATTAATAATTTTGGAACTGTGAAGATTAATA 2100  
Db 5781 TGGCTCACTTCATGATGATTAATAATGATTAATAATTTTGGAACTGTGAAGATTAATA 5840  
Qy 2101 TACCAATAATAATATATATAAAAGTATTTATATGAAATTAATAATAAATCAGATGAT 2160  
Db 5841 TACCAATAATAATATATATAAAAGTATTTATATGAAATTAATAATAAATCAGATGAT 5900  
Qy 2161 GGAATAAAGTTG 2172  
Db 5901 GGAATAAAGTTG 5912

RESULT 5  
BD016860  
LOCUS BD016860 11204 bp DNA linear PART 27-AUG-2002  
DEFINITION Novel cytidine deaminase.  
ACCESSION BD016860  
VERSION BD016860.1 GI:22558036  
KEYWORDS JP 2001245669-A/33.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Molecular Function: Cytidine deaminase activity  
Biological Process: Nucleotide metabolism  
Cellular Component: Cytoplasm  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 11204)  
AUTHORS Honjo, T. and Muramatsu, M.  
TITLE Novel cytidine deaminase  
JOURNAL Patent: JP 2001245669-A 33 11-SEP-2001;  
JAPAN TOBACCO INC, TASUKU HONJO  
COMMENT OS Homo sapiens (human)  
PD JP 2001245669-A/33  
PF 11-SEP-2001  
PI 28-MAR-2000 JP 2000092981  
PI TASUKU HONJO, MASAMICHI MURAMATSU  
PC C12N15/09, A61K39/395, A61K39/395, A61P1/00, A61P11/06, A61P13/12,  
PC A61P17/00  
PC A61P27/02, A61P27/16, A61P37/02, A61P37/08, C07K16/18, C12N1/19, PC  
C12N1/21,  
PC C12N5/10, C12N9/78, C12P21/02, C12P21/08// (C12N1/21, C12R1/19), PC  
(C12N5/10, C12R1/91), C12N15/00, C12N5/00, (C12N5/00, C12R1/91) CC  
FH Key Location/Qualifiers  
1. 11204  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

ORIGIN  
Query Match 100.0%; Score 2172; DB 6; Length 11204;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCCCCGTATGAGGTGATGATGATTAAGAGCGCATTTCTGACTTTGGGACTTTGATAGCA 60  
Db 8956 CCCCCTATGAGGTGATGATGATTAAGAGCGCATTTCTGACTTTGGGACTTTGATAGCA 9015

Qy 61 CTTCCAGAAATGTCACACAGATGAATATCTCTGCTGAAGACAGTGAATAAAAAACGT 120  
Db 9016 CTTCCAGAAATGTCACACAGATGAATATCTCTGCTGAAGACAGTGAATAAAAAACGT 9075  
Qy 121 CTTCAAGCTCTCTCTGTTTTTATCTCTCACTCTCACTTTCTTGAAGTTTACAGAAAA 180  
Db 9076 CTTCAAGCTCTCTCTGTTTTTATCTCTCACTCTCACTTTCTTGAAGTTTACAGAAAA 9135  
Qy 181 ATATTTATATACAGCTTTTAAAAAAGATCTATGCTTTGAAAAATAGAGAAAGAACAGAG 240  
Db 9136 ATATTTATATACAGCTTTTAAAAAAGATCTATGCTTTGAAAAATAGAGAAAGAACAGAG 9195  
Qy 241 CTGCGCAGGAGCGTGCAATGGTGCAGTTTGTGAATGCAACATTTGCCCTACTGGGA 300  
Db 9196 CTGCGCAGGAGCGTGCAATGGTGCAGTTTGTGAATGCAACATTTGCCCTACTGGGA 9255  
Qy 301 ATACAGAACTGACAGAACTGGAGCATCTTAAAGTGTCAACGTTTTCTATGACTTTTA 360  
Db 9256 ATACAGAACTGACAGAACTGGAGCATCTTAAAGTGTCAACGTTTTCTATGACTTTTA 9315  
Qy 361 GGTAGATGAGAGAGAGAGGTATCTTAAAGATGTGAGAGAGATCAATGTTTTTA 420  
Db 9316 GGTAGATGAGAGAGAGAGGTATCTTAAAGATGTGAGAGAGATCAATGTTTTTA 9375  
Qy 421 TATCAACATCCTTATATTTGATTCATTTGATTAAGTTAAAGTGTGATGATAGATTTT 480  
Db 9376 TATCAACATCCTTATATTTGATTCATTTGATTAAGTTAAAGTGTGATGATAGATTTT 9435  
Qy 481 TCTATCTTTCCCTTACAGTTACTTTCAAGTAAACAACTCTTCAATCAGGCCATGA 540  
Db 9436 TCTATCTTTCCCTTACAGTTACTTTCAAGTAAACAACTCTTCAATCAGGCCATGA 5495  
Qy 541 TCTATGAGCTCTCTAATGAGATATCTGGGTATTTGACCCCAACATCTTCCAAA 600  
Db 9496 TCTATGAGCTCTCTAATGAGATATCTGGGTATTTGACCCCAACATCTTCCAAA 9555  
Qy 601 GCATTAATATCCATCATGCTGCTGATGTTTATATAGAGAGAGATGTTTTATGTTTG 660  
Db 9556 GCATTAATATCCATCATGCTGCTGATGTTTATATAGAGAGAGATGTTTTATGTTTG 9615  
Qy 661 TACAAAGAAAGATTTGATGAGTGGAGTGAAGTATGACATGATGATGATGATGATGAT 720  
Db 9616 TACAAAGAAAGATTTGATGAGTGGAGTGAAGTATGACATGATGATGATGATGATGAT 9675  
Qy 721 GCTACTTTAATAAGATCTTAAATGGGAGAGAGATGTGAACAAGACACCTTATTA 780  
Db 9676 GCTACTTTAATAAGATCTTAAATGGGAGAGAGATGTGAACAAGACACCTTATTA 9735  
Qy 781 TGGGTGATGCTGGAAGTGAAGAAATCTTGTGAAGCGAACTCTTTTAAAGAAAGTCCCT 840  
Db 9736 TGGGTGATGCTGGAAGTGAAGAAATCTTGTGAAGCGAACTCTTTTAAAGAAAGTCCCT 9795  
Qy 841 AATTATGAACACCCCAAACTTCAATATCAATATTAAGCAAACTTTGAAGAAAGTGG 900  
Db 9796 AATTATGAACACCCCAAACTTCAATATCAATATTAAGCAAACTTTGAAGAAAGTGG 9855  
Qy 901 CTTGAATGTTGGGAGAGAAATCTATGAGCTCTGCTGAGTCTTCAATCTCAGAAATG 960  
Db 9856 CTTGAATGTTGGGAGAGAAATCTATGAGCTCTGCTGAGTCTTCAATCTCAGAAATG 9915  
Qy 961 CCAATCAGTCAAGGTTTGTCTACATTTGTATGTGTATGCTTCTCCAAAGATAT 1020  
Db 9916 CCAATCAGTCAAGGTTTGTCTACATTTGTATGTGTATGCTTCTCCAAAGATAT 9975  
Qy 1021 TAACTATATAGAGATTTGACAAAACAGATGATTAAGCTTCCGAACTGTGCAACGC 1080  
Db 9976 TAACTATATAGAGATTTGACAAAACAGATGATTAAGCTTCCGAACTGTGCAACGC 10035  
Qy 1081 TCAATGTTCAAGCTGCTGGAGGTTGAGAGGAGAGATGAGCTTGAACAAGGTGTTCAA 1140  
Db 10036 TCAATGTTCAAGCTGCTGGAGGTTGAGAGGAGAGATGAGCTTGAACAAGGTGTTCAA 10095



QY 1141 GGCCAGCCTGGGCAACATACAGAGATCCTGCTCTCAAAAAAAAAAAAAAAAAAGAAA 1200  
 Db 10096 GGCCAGCCTGGGCAACATACAGAGATCCTGCTCTCAAAAAAAAAAAAAAAAAAGAAA 10155  
 QY 1201 GAGAGAGGCGCGCGCTGCTGCTCAAGCCTGTATCCAGAGACTTTGGAGCGCGAGCC 1260  
 Db 10156 GAGAGAGGCGCGCGCTGCTGCTCAAGCCTGTATCCAGAGACTTTGGAGCGCGAGCC 10215  
 QY 1261 GGGCGGATCACTGTGTGTGAGAGAGTTTGAACAGCGCTGGCCAAATGGCAAAACCCCGT 1320  
 Db 10216 GGGCGGATCACTGTGTGTGAGAGAGTTTGAACAGCGCTGGCCAAATGGCAAAACCCCGT 10275  
 QY 1321 CTGTACTCAAAATGCAAAATATGACGAGCGTGTGTAGAGAGAGCTGTATCCAGCTAC 1380  
 Db 10276 CTGTACTCAAAATGCAAAATATGACGAGCGTGTGTAGAGAGAGCTGTATCCAGCTAC 10335  
 QY 1381 TTGGAGGCTGAGGAGAGAGAAATCGCTTGAACCCAGAGGTGAGGTTGCACTAAGCTGA 1440  
 Db 10336 TTGGAGGCTGAGGAGAGAGAAATCGCTTGAACCCAGAGGTGAGGTTGCACTAAGCTGA 10395  
 QY 1441 GATCGTGGCTTGCATCTCCAGCTGGCGCAAGAGAGAGACTCTGTCTCAAAAAAAA 1500  
 Db 10396 GATCGTGGCTTGCATCTCCAGCTGGCGCAAGAGAGAGACTCTGTCTCAAAAAAAA 10455  
 QY 1501 AAAAAAGAT 1560  
 Db 10456 AAAAAAGAT 10515  
 QY 1561 TGCAGAGAAATGTGCTTTATCCAGAAATGTAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1620  
 Db 10516 TGCAGAGAAATGTGCTTTATCCAGAAATGTAGAGAGAGAGAGAGAGAGAGAGAGAT 10575  
 QY 1621 TCTCTTTTGTGTCTATTTTGTCCCTAACAAGTGTCTTTTGAAGAGAGAGAGAGAGAT 1680  
 Db 10576 TCTCTTTTGTGTCTATTTTGTCCCTAACAAGTGTCTTTTGAAGAGAGAGAGAGAT 10635  
 QY 1681 ATACCAATCCCTGTGCGGTATTTACCTAGCAAGCTTGCATGAGAGAGAGAGAGAT 1740  
 Db 10636 ATACCAATCCCTGTGCGGTATTTACCTAGCAAGCTTGCATGAGAGAGAGAGAT 10695  
 QY 1741 ACAGAGAACTGAATGCAACAAGTCTTATTTATTTATTTATTTATTTATTTATTTATTT 1800  
 Db 10696 ACAGAGAACTGAATGCAACAAGTCTTATTTATTTATTTATTTATTTATTTATTTATTT 10755  
 QY 1801 AGAGTTAAAAATGTATTCATGATTCATTTATTTATTTATTTATTTATTTATTTATTTAT 1860  
 Db 10756 AGAGTTAAAAATGTATTCATGATTCATTTATTTATTTATTTATTTATTTATTTATTTAT 10815  
 QY 1861 ATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1920  
 Db 10816 ATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 10875  
 QY 1921 AATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1980  
 Db 10876 AATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 10935  
 QY 1981 GGTGCTACGAGAGCAATTTCTGTGATTTTATTTATTTATTTATTTATTTATTTATTTAT 2040  
 Db 10936 GGTGCTACGAGAGCAATTTCTGTGATTTTATTTATTTATTTATTTATTTATTTATTTAT 10995  
 QY 2041 TGGCTCATTTCAATCATGTTAAATTAATGATAAATTTTGAAGCTGTGAAGATTA 2100  
 Db 10996 TGGCTCATTTCAATCATGTTAAATTAATGATAAATTTTGAAGCTGTGAAGATTA 11055  
 QY 2101 TACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2160  
 Db 11056 TACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 11115  
 QY 2161 GGAATTAAGTTG 2172  
 Db 11116 GGAATTAAGTTG 11127

RESULT 6  
 AB040430  
 LOCUS  
 DEFINITION  
 Homo sapiens AID gene for activation-induced cytidine deaminase,  
 complete cds.  
 ACCESSION  
 AB040430.1  
 VERSION  
 AB040430.1  
 KEYWORDS  
 AID; activation-induced cytidine deaminase.  
 SOURCE  
 Homo sapiens (human)  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 1 (sites)  
 Muto,T., Muramatsu,M., Tanikawa,M., Kinoshita,K. and Honjo,T.  
 Isolation, tissue distribution, and chromosomal localization of the  
 human activation-induced cytidine deaminase (AID) gene  
 Genomics 68 (1), 85-88 (2000)  
 MEDLINE  
 20408890  
 PUBMED  
 10950930  
 REFERENCE  
 2 (sites)  
 Revy,P., Muto,T., Levy,Y., Geisemann,F., Plehant,A., Sanaj,O.,  
 Catalan,M., Forveille,M., Dufourcq-Iagelouse,R., Gennery,A.,  
 Tazcan,I., Brody,F., Kayserili,H., Ugazio,A.G., Brousee,N.,  
 Muramatsu,M., Notarangelo,L.D., Kinoshita,K., Honjo,T., Fischer,A.  
 and Durandy,A.  
 Activation-induced cytidine deaminase (AID) deficiency causes the  
 autosomal recessive form of the Hyper-IgM syndrome (HIGM2)  
 Cell 102 (5), 565-575 (2000)  
 MEDLINE  
 20460541  
 PUBMED  
 1107475  
 REFERENCE  
 3 (bases 1 to 11204)  
 Muto,T., Muramatsu,M., Tanikawa,M., Kinoshita,K. and Honjo,T.  
 Direct Submission  
 Submitted (18-MAR-2000) Tasuku Honjo, Kyoto University, Department  
 of Medical Chemistry, Faculty of Medicine, Yoshida, Sakyo-ku,  
 Kyoto, Kyoto 606-8501, Japan (E-mail: honjo@omf.kyoto-u.ac.jp,  
 Tel:81-75-753-4371 (ex. 4371), Fax:81-75-753-4388)  
 Location/Qualifiers  
 source  
 1. 11204  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 join(521..528,6280..6427,7807..8077,8371..8486,8956..9009)  
 /gene="AID"  
 join(521..528,6280..6427,7807..8077,8371..8486,8956..9009)  
 /gene="AID"  
 /codon\_start=1  
 /product="activation-induced cytidine deaminase"  
 /protein\_id="BA012720.1"  
 /db\_xref="GI:9988408"  
 /translation="MDSLMMRRKFLYQKRWAKRRETYCIYVKRDSATSL  
 DFGYLRNMGCHVELFLRYISDWDLPGRVFTWFTSPCYDCAHVAADFLRNP  
 NLSRIETALYFCEDRKASPEGLRLHAGVLIAMTFEDYFCWNTFEVNERTEK  
 AMEGLHENSRLRQLRILLPLRYVDLRLDAERTLG"  
 ORIGIN  
 Query Match 100.0%; Score 2172; DB 9; Length 11204;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CCCTGTATGAGTTGATGACTTTCGAGAGAGCAATTCCTTTGGAGCTTTATAGCAA 60.  
 Db 8956 CCCCTGTATGAGTTGATGACTTTCGAGAGAGCAATTCCTTTGGAGCTTTATAGCAA 9015  
 QY 61 CTTCAGGAATGACACAGATGAATATCTCTGCTGAAGACAGTGAATAAAAACAGT 120  
 Db 9016 CTTCAGGAATGACACAGATGAATATCTCTGCTGAAGACAGTGAATAAAAACAGT 9075  
 QY 121 CCTCAAGCTTCTCTGTTTATTTCTTCAACTCTCACTTTCTTAGAGTTTACAGAAAA 180  
 Db 9076 CCTCAAGCTTCTCTGTTTATTTCTTCAACTCTCACTTTCTTAGAGTTTACAGAAAA 9135  
 QY 181 ATATTATATACGACTTTTAAAAAGATCTATGCTTTGAAAAATAGAGAGAAACAGGT 240



Db 9136 AATATTATATAGACACTTTTAAAAAGATCTATGTCTTGAATAATAGAGAAACAAGT 9195  
Qy 241 CTGGCAGGAGAGCTGTGCAATTTGGGAGTTTGTGAATGCAACATTTGCCCTATGAGGA 300  
Db 9196 CTGGCAGGAGAGCTGTGCAATTTGGGAGTTTGTGAATGCAACATTTGCCCTATGAGGA 9255  
Qy 301 ATAGAGAACTGACAGAGCTGGAGATCTTAAAGTGTCAAGTCTTCTATGACTTTTA 360  
Db 9256 ATAGAGAACTGACAGAGCTGGAGATCTTAAAGTGTCAAGTCTTCTATGACTTTTA 9315  
Qy 361 GGTAGAGTGAAGCAGAGAGTGTATCTTAAAGATGGTGAAGAGATCAATGTTTTA 420  
Db 9316 GGTAGAGTGAAGCAGAGAGTGTATCTTAAAGATGGTGAAGAGATCAATGTTTTA 9375  
Qy 421 TATCAACATCTTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 480  
Db 9376 TATCAACATCTTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 9435  
Qy 481 TCTATCTTTTCCCTTGAAGTTTATCTTCAAGTAAACAACCTCTTCATCAGGCAATGA 540  
Db 9436 TCTATCTTTTCCCTTGAAGTTTATCTTCAAGTAAACAACCTCTTCATCAGGCAATGA 9495  
Qy 541 TCTATAGACCTCTCATAGAGATCTGGGTGATGTGACCCCAACATCTCTCCAA 600  
Db 9496 TCTATAGACCTCTCATAGAGATCTGGGTGATGTGACCCCAACATCTCTCCAA 9555  
Qy 601 GCATTAATATCCAAATCATGCTGTATGTTTAAATCAAGAGAGATGTTTATGTTTG 660  
Db 9556 GCATTAATATCCAAATCATGCTGTATGTTTAAATCAAGAGAGATGTTTATGTTTG 9615  
Qy 661 TCAAAAAGAGATGTTTATGGGTGGAGATGAGATGATGACATGATGATGACCTTGA 720  
Db 9616 TCAAAAAGAGATGTTTATGGGTGGAGATGAGATGATGACATGATGATGACCTTGA 9675  
Qy 721 GCTACTTTATTAAGATCTTAAATGGGAGAGAGATGATGATGATGATGATGATGAT 780  
Db 9676 GCTACTTTATTAAGATCTTAAATGGGAGAGAGATGATGATGATGATGATGATGAT 9735  
Qy 781 TGGGTGATGTGTGAAGTGAAGTCTTGTGAAAAGCAACTCTTTTAAAGAGTCCCT 840  
Db 9736 TGGGTGATGTGTGAAGTGAAGTCTTGTGAAAAGCAACTCTTTTAAAGAGTCCCT 9795  
Qy 841 AATTTTGAAGACCCCAAACTTCAATATCAATTTAGCAAAATGGAAGAGTTG 900  
Db 9796 AATTTTGAAGACCCCAAACTTCAATATCAATTTAGCAAAATGGAAGAGTTG 9855  
Qy 901 CTGGAATGTTGGGAGAGAGAAATCTATGAGTCTGAGGCTCTTCAATCAGAAATG 960  
Db 9856 CTGGAATGTTGGGAGAGAGAAATCTATGAGTCTGAGGCTCTTCAATCAGAAATG 9915  
Qy 961 CCAATCAGGTCAAGTTTGTCTACATTTTGTATGTTGATGATCTTCCCAAGTATAT 1020  
Db 9916 CCAATCAGGTCAAGTTTGTCTACATTTTGTATGTTGATGATCTTCCCAAGTATAT 9975  
Qy 1021 TAACTATATAGAGAGTGTGACAAACAGATGATTAAGTGTGCAACCGTGGACACGC 1080  
Db 9976 TAACTATATAGAGAGTGTGACAAACAGATGATTAAGTGTGCAACCGTGGACACGC 10035  
Qy 1081 TCAATGTTCTAGTGTCTGGAGAGTTGAGAGAGAGAGATGCTTGAACACAGTGTCAA 1140  
Db 10036 TCAATGTTCTAGTGTCTGGAGAGTTGAGAGAGAGAGATGCTTGAACACAGTGTCAA 10095  
Qy 1141 GGCAGAGCTGGGCAACATAAGATCTGTCTCAAAAAAATTTTAAAAAAGAAA 1200  
Db 10096 GGCAGAGCTGGGCAACATAAGATCTGTCTCAAAAAAATTTTAAAAAAGAAA 10155  
Qy 1201 GAGAGAGGCGCGGCGTGTGTGCTCAGCGCTGTATCCAGAGACTTTGGGAGCGGAGCC 1260  
Db 10156 GAGAGAGGCGCGGCGTGTGTGCTCAGCGCTGTATCCAGAGACTTTGGGAGCGGAGCC 10215  
Qy 1261 GGCAGAGTCACTGTGTGCTCAGAGTGTGAGACCAAGCTTGGCCCAATGCGCAAAACCCCGT 1320

Db 10216 GGCAGAGTCACTGTGTGCTCAGAGTGTGAGACCAAGCTTGGCCCAATGCGCAAAACCCCGT 10275  
Qy 1321 CTGACTCAAAATATGCAAAATTTAGCAGGCGGTGTAGAGGACCACTGTATCCCACTAC 1380  
Db 10276 CTGACTCAAAATATGCAAAATTTAGCAGGCGGTGTAGAGGACCACTGTATCCCACTAC 10335  
Qy 1381 TTGGAGGCTGAGGACAGAGAAATGCTTGAACCCAGAGGTGAGAGTTGCAATGACTGA 1440  
Db 10336 TTGGAGGCTGAGGACAGAGAAATGCTTGAACCCAGAGGTGAGAGTTGCAATGACTGA 10395  
Qy 1441 GATGTGCGGTGTGCTCAGCGCTGGGCGCAAGACCAAGACTTGTCTCAGAAAAAAA 1500  
Db 10396 GATGTGCGGTGTGCTCAGCGCTGGGCGCAAGACCAAGACTTGTCTCAGAAAAAAA 10455  
Qy 1501 AAAAAAAGAT 1560  
Db 10456 AAAAAAAGAT 10515  
Qy 1561 TGCAGAGAAATTTGCTTTATCCCAAAATGTAGAGAGCAATTAAGGATCCCTATTTG 1620  
Db 10516 TGCAGAGAAATTTGCTTTATCCCAAAATGTAGAGAGCAATTAAGGATCCCTATTTG 10575  
Qy 1621 TCTCTTTGGTGTCTATTTGTCTCTTAAACAAGTCTTTGAAGTGAAGAAATATTCAGA 1680  
Db 10576 TCTCTTTGGTGTCTATTTGTCTCTTAAACAAGTCTTTGAAGTGAAGAAATATTCAGA 10635  
Qy 1681 ATTAACCATATCCCTGTGCGCTTATTAACCTAGCAACCTTGCATTAAGATGAGAGATCC 1740  
Db 10636 ATTAACCATATCCCTGTGCGCTTATTAACCTAGCAACCTTGCATTAAGATGAGAGATCC 10695  
Qy 1741 ACAAGAAATCTGAATGACAACTGTCTTATTAATCTTATGATGATGATGATGATGAT 1800  
Db 10696 ACAAGAAATCTGAATGACAACTGTCTTATTAATCTTATGATGATGATGATGATGAT 10755  
Qy 1801 AGAGTTAAATTTGATCTTATGATGATGATGATGATGATGATGATGATGATGATGATG 1860  
Db 10756 AGAGTTAAATTTGATCTTATGATGATGATGATGATGATGATGATGATGATGATGATG 10815  
Qy 1861 AATTTTATTAACATGATTTCTTTCTGTATGATGATGATGATGATGATGATGATGATGAT 1920  
Db 10816 AATTTTATTAACATGATTTCTTTCTGTATGATGATGATGATGATGATGATGATGATGAT 10875  
Qy 1921 AATTTTAACTTGAAGATGATCTTAATTAACAAGTATGATGATGATGATGATGATGAT 1980  
Db 10876 AATTTTAACTTGAAGATGATCTTAATTAACAAGTATGATGATGATGATGATGATGAT 10935  
Qy 1981 GGTGCTACAGAGCAATTTCTGATTTTATGAACTTTATGACAGCAAAATTTGCTTC 2040  
Db 10936 GGTGCTACAGAGCAATTTCTGATTTTATGAACTTTTATGACAGCAAAATTTGCTTC 10995  
Qy 2041 TGGCTCACTTCAATCAGTTAAATTAATGATTAATTTTGAAGCTGTGAAGATPAAA 2100  
Db 10996 TGGCTCACTTCAATCAGTTAAATTAATGATTAATTTTGAAGCTGTGAAGATPAAA 11055  
Qy 2101 TACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2160  
Db 11056 TACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 11115  
Qy 2161 GGAATTAACCTG 2172  
Db 11116 GGAATTAACCTG 11127

RESULT 7  
AC092184 71132 bp DNA linear PRI 12-JUN-2002  
LOCUS AC092184  
DEFINITION Homo sapiens 12 BAC RP11-438b7 (Roswell Park Cancer Institute Human  
BAC library) complete sequence.  
ACCESSION AC092184 AC013443  
VERSION AC092184.7 GI:21206067  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens



REFERENCE  
AUTHORS

Eukaryotes; Metazoa; Chordata; Cranata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Carnivora; Hominoidea; Homo.  
1 (bases 1 to 71132)

Muzny, D.M., Adams, C., Adio-Ogunola, B., Ali-Osman, F.R., Allen, C.,



```
repeat_region      5305. .5563      /rpt_family="AluYo"
repeat_region      5564. .5599      /rpt_family="AT_rich"
repeat_region      5644. .5733      /rpt_family="MSTD"
repeat_region      5734. .6012      /rpt_family="AluSg"
repeat_region      6020. .6336      /rpt_family="AluSx"
repeat_region      6337. .6629      /rpt_family="MSTD"
repeat_region      6630. .6650      /rpt_family="(CA)n"
repeat_region      6711. .6817      /rpt_family="MIR"
repeat_region      7033. .7327      /rpt_family="AluSx"
repeat_region      7607. .7745      /rpt_family="FLAM_A"
repeat_region      7980. .8047      /rpt_family="MIR"
repeat_region      8050. .8549      /rpt_family="LTR47A"
repeat_region      8567. .8698      /rpt_family="MTR"
repeat_region      8815. .9101      /rpt_family="AluSg"
repeat_region      11227. .12175      /rpt_family="LTR5"
STS
12625. .12900      /standard_name="87432"
repeat_region      13547. .13664      /rpt_family="LIMB5"
repeat_region      13665. .13979      /rpt_family="AluY"
repeat_region      13980. .14193      /rpt_family="LIMB5"
repeat_region      14622. .14924      /rpt_family="AluY"
repeat_region      15813. .17336      /rpt_family="LIPa8"
repeat_region      17272. .17395      /rpt_family="FLAM_C"
repeat_region      17426. .17575      /rpt_family="MIR"
```

Query Match	100.0%	Score 2172	DB 9	Length 71132
Best Local Similarity	100.0%	Pred. No. 0		
Matches 2172	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1	CCCCGTTATGAGCTTGATGACTTACGAAACGATTTGCTACTTTGGGACTTTGATAGCA	60	
Db	43499	CCCCGTTATGAGCTTGATGACTTACGAAACGATTTGCTACTTTGGGACTTTGATAGCA	43555	
QY	61	CTTCCAGGAATGTCAACACAGGTGAATATCTGCTGTAAGACAGGTGATAAAAACGT	120	
Db	43559	CTTCCAGGAATGTCAACACAGGTGAATATCTGCTGTAAGACAGGTGATAAAAACGT	43615	
QY	121	CCTTCAAGTCTTCTGTTTTTATCTTCAACTCTCACTTCTTACGATTACAGAAAA	180	
Db	43619	CCTTCAAGTCTTCTGTTTTTATCTTCAACTCTCACTTCTTACGATTACAGAAAA	43675	
QY	181	ATATTATATACGACTCTTTAAAAAGATCTATGCTTGAAAAATGAGAGACACAGCT	240	
Db	43679	ATATTATATACGACTCTTTAAAAAGATCTATGCTTGAAAAATGAGAGACACAGCT	43735	
QY	241	CTGGCCAGGAGCGTCTGCAATTTGTGTGAGTTTGATGACAACTGTGCCCTACTGGGA	300	
Db	43739	CTGGCCAGGAGCGTCTGCAATTTGTGTGAGTTTGATGACAACTGTGCCCTACTGGGA	43795	
QY	301	ATTACAGACTGACGAGACCTGGGAGCACTTAAAGTGCAACGTTTTCTATGACCTTTA	360	

Db	43789	ATAACAAACTGCAAGACCTGGAGCATCTCTAAAGTCAACGCTTTTCTATGACTTTA	43885
Qy	361	GGTAGATGAGAGCAGAGGTAGATCCTATAAAGCATGCTGAGAGATCAATGTTTTA	420
Db	43859	GGTAGATGAGAGCAGAGGTAGATCCTATAAAGCATGCTGAGAGATCAATGTTTTA	43918
Qy	421	TATCAACATCCTTATATTTGATTCATTTGAATTAAAGTGGTGTAGATAGATTTT	480
Db	43919	TATCAACATCCTTATATTTGATTCATTTGAATTAAAGTGGTGTAGATAGATTTT	43978
Qy	481	TCATATCTTTTCCCTGAGCGTTACTTCAAGTACAAACAACTCTTCCATCAGGCGCATGA	540
Db	43979	TCATATCTTTTCCCTGAGCGTTACTTCAAGTACAAACAACTCTTCCATCAGGCGCATGA	44038
Qy	541	TCATAGAGACCTCCTAATGAGATATCGGTGATTTGTGACCCCAACCATCTTCCAA	600
Db	44039	TCATAGAGACCTCCTAATGAGATATCGGTGATTTGTGACCCCAACCATCTTCCAA	44098
Qy	601	GCATTATATCCATCATGCGCTGATGTTTATACGACGAAGCATGTTTTATGTG	660
Db	44099	GCATTATATCCATCATGCGCTGATGTTTATACGACGAAGCATGTTTTATGTG	44158
Qy	661	TACAAAGAGATTTGTATGGGTGGGGATGGAGTTPACACCATGATGCTCA	720
Db	44159	TACAAAGAGATTTGTATGGGTGGGGATGGAGTTPACACCATGATGCTCA	44218
Qy	721	GCTACTTTAATTAAGATCTTAAATATGGGACAGAGACTGTGACAAACACCTTATA	780
Db	44219	GCTACTTTAATTAAGATCTTAAATATGGGACAGAGACTGTGACAAACACCTTATA	44278
Qy	781	TGGTGTATGCTCTGAAGTACAAATCTTCTGAAACGCAACTCTTTAAGAAATCCT	840
Db	44279	TGGTGTATGCTCTGAAGTACAAATCTTCTGAAACGCAACTCTTTAAGAAATCCT	44338
Qy	841	AATTTAAGAACCCCAACCTTCACATATCAATATTTAGCAACATTTGGAGGAAGTTG	900
Db	44339	AATTTAAGAACCCCAACCTTCACATATCAATATTTAGCAACATTTGGAGGAAGTTG	44398
Qy	901	CTTAATAGTTGGGGAGAGAAAATCTATTTGGCTCTGTGGGTCTCTTCACTCAGAAATG	960
Db	44399	CTTAATAGTTGGGGAGAGAAAATCTATTTGGCTCTGTGGGTCTCTTCACTCAGAAATG	44458
Qy	961	CCAAATCGGTCAGAGTTTGTACATTTTGTATGTGTATGTGATGCTTCTCCAAAGTATAT	1020
Db	44459	CCAAATCGGTCAGAGTTTGTACATTTTGTATGTGTATGTGATGCTTCTCCAAAGTATAT	44518
Qy	1021	TAACTATATAGAGAGTTTGTACAAACAGATATTAAGCTGCCAACCTGTGGCACACGC	1080
Db	44519	TAACTATATAGAGAGTTTGTACAAACAGATATTAAGCTGCCAACCTGTGGCACACGC	44578
Qy	1081	TCATAGTTCTAGCTGCTGTGGAGGTTTGAAGAGGAGAGATGCGCTTGAAACAGAGTTTCA	1140
Db	44579	TCATAGTTCTAGCTGCTGTGGAGGTTTGAAGAGGAGAGATGCGCTTGAAACAGAGTTTCA	44638
Qy	1141	GGCCAGCCTGGGCACATTAACAAGATCTGTCTCTCAAAAAAAAAAAAAAAAAAAAAA	1200
Db	44639	GGCCAGCCTGGGCACATTAACAAGATCTGTCTCTCAAAAAAAAAAAAAAAAAAAAAA	44698
Qy	1201	GAGAGAGGAGCCGGGCGTGTGTGCTCACGCTGTATATCCACGACTTTGGGAGCGGAGCC	1260
Db	44699	GAGAGAGGAGCCGGGCGTGTGTGCTCACGCTGTATATCCACGACTTTGGGAGCGGAGCC	44758
Qy	1261	GGGGGATATCACGTGTGGTCAGAGGTTTGAACAAGCTGGCCCAACAATGGCAAAACCCCGT	1320
Db	44759	GGGGGATATCACGTGTGGTCAGAGGTTTGAACAAGCTGGCCCAACAATGGCAAAACCCCGT	44818
Qy	1321	CTGTACTCAAAATGCAAAATTTAGCCAGGCGTGTAGCAGGACCTGTATCCAGCTTAC	1380
Db	44819	CTGTACTCAAAATGCAAAATTTAGCCAGGCGTGTAGCAGGACCTGTATCCAGCTTAC	44878
Qy	1381	TTGGAGAGGCTGAGGACGAGAGATTCGTTTAACCCAGAGAGTGTGAGGTTTCAATAGCTGA	1440
Db	44879	TTGGAGAGGCTGAGGACGAGAGATTCGTTTAAACCCAGAGAGTGTGAGGTTTCAATAGCTGA	44938







## ORIGIN

NI5LR1FTARLYPCEDRKAPEBGLRLRHRAGVQIAIMTFKQFYCWMTFVENHERTEK  
AMEGHENSVRSLRQLRRLILPEYVDLDRFAETLGL"

Query Match 55.2%; Score 1199.6; DB 9; Length 1828;  
Best Local Similarity 99.7%; Pred. No. 3.5e-221;  
Matches 1202; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 CCCGCTGATAGAGTGTGATGACTTACGAGACGATTTGCTGACTTTGGGACTTTGATAGCA 60  
DB 620 CCCCTGTATGAGAGTGTGATGACTTACGAGACGATTTGCTGACTTTGGGACTTTGATAGCA 679  
QY 61 CTTCAGAGATGTCAACACGATGAAATATCTGCTGTAAGACAGTGTGATAAAAAAGT 120  
DB 680 CTTCAGAGATGTCAACACGATGAAATATCTGCTGTAAGACAGTGTGATAAAAAAGT 739  
QY 121 CCTTCAAGTCTTCTGTTTTTATCTTAACTTCACTTCTTGAAGTTTACAGAAAA 180  
DB 740 CCTTCAAGTCTTCTGTTTTTATCTTCACTTCTTGAAGTTTACAGAAAA 799  
QY 181 ATATTATATAGAGCTCTTAAAGATCTATGCTGTAAGTAAGAAAGACAGAGT 240  
DB 800 ATATTATATAGAGCTCTTAAAGATCTATGCTGTAAGTAAGAAAGACAGAGT 859  
QY 241 CTGGCCAGAGAGTGTGCTGCAATTTGTCAGTTTGAATGCAACTTGTCCCTACTGGA 300  
DB 860 CTGGCCAGAGAGTGTGCTGCAATTTGTCAGTTTGAATGCAACTTGTCCCTACTGGA 919  
QY 301 ATTAACAGAACTGACAGACCTGGAGACCTTAAGTGTCAAGTTTCTATGACTTTTA 360  
DB 920 ATTAACAGAACTGACAGACCTGGAGACCTTAAGTGTCAAGTTTCTATGACTTTTA 979  
QY 361 GGTAGATATAGACAGAGATGATCTTAAAGCATGCTGAGAGATCAATGTTTTTA 420  
DB 980 GGTAGATATAGACAGAGATGATCTTAAAGCATGCTGAGAGATCAATGTTTTTA 1039  
QY 421 TATCAACATCCTTATATTTATGATGATGATGATGATGATGATGATGATGATGATGAT 480  
DB 1040 TATCAACATCCTTATATTTATGATGATGATGATGATGATGATGATGATGATGATGAT 1099  
QY 481 TCTATCTTTTCCCTTGAAGTTTACTTTCAAGTAACAACTCTTCAATCAGGAGATGA 540  
DB 1100 TCTATCTTTTCCCTTGAAGTTTACTTTCAAGTAACAACTCTTCAATCAGGAGATGA 1159  
QY 541 TCTATAGACCTCTTATAGAGATCTGGGATGATGATGATGATGATGATGATGATGATGAT 600  
DB 1160 TCTATAGACCTCTTATAGAGATCTGGGATGATGATGATGATGATGATGATGATGATGAT 1219  
QY 601 GCATTTATATCCATATCATGCTGTATGTTTTATATCAGAGAAAGATTTTTTATGTTTG 660  
DB 1220 GCATTTATATCCATATCATGCTGTATGTTTTATATCAGAGAAAGATTTTTTATGTTTG 1279  
QY 661 TACAAAGAGAGATGTTATGGGAGGAGATGAGATATAGACATGATGCTACCTTCAA 720  
DB 1280 TACAAAGAGAGATGTTATGGGAGGAGATGAGATATAGACATGATGCTACCTTCAA 1339  
QY 721 GCTACTTTATTAAGAGATCTTAAATGGGACAGAGACTGTGACAAACACCTTATTA 780  
DB 1340 GCTACTTTATTAAGAGATCTTAAATGGGACAGAGACTGTGACAAACACCTTATTA 1399  
QY 781 TGGGTTGATGTCTGAAGTACAAATCTTTGAAAACGCAACTCTTTTAAAGAGTCCCT 840  
DB 1400 TGGGTTGATGTCTGAAGTACAAATCTTTGAAAACGCAACTCTTTTAAAGAGTCCCT 1459  
QY 841 AATTATAGAAACACCAACCACTTCAATATCAATTAATAGCAACATTTGAAAGAGTTG 900  
DB 1460 AATTATAGAAACACCAACCACTTCAATATCAATTAATAGCAACATTTGAAAGAGTTG 1519  
QY 901 CTGTAATGTGGGAGAGAGAAATCTATTGGCTCTGCTGCTCTTCAATCTCAGAAATG 960  
DB 1520 CTGTAATGTGGGAGAGAGAAATCTATTGGCTCTGCTGCTCTTCAATCTCAGAAATG 1579  
QY 961 CCAATCAGGTCAAGTTTGCTAATTTTGTATGTGTGATGCTTCCCAAGGTATAT 1020

DB 1580 CCAATCAGGTCAAGTTTGCTAATTTTGTATGTGTGATGCTTCCCAAGGTATAT 1639  
QY 1021 TAATATATAGAGAGTGTGACAAACAGATATTAAGCTGCGAAGCTGACACAGC 1080  
DB 1640 TAATATATAGAGAGTGTGACAAACAGATATTAAGCTGCGAAGCTGACACAGC 1699  
QY 1081 TCAATAGTTTACGCTGTTGGAGAGTTGAGAGAGAGATGCTTGAACACAGGTGTCAA 1140  
DB 1700 TCAATAGTTTACGCTGTTGGAGAGTTGAGAGAGAGATGCTTGAACACAGGTGTCAA 1759  
QY 1141 GGCACGCTGGGACATTAACAAGATCTGTCTCTCAAAAAAAAAAAAAAAAAAAAAA 1200  
DB 1760 GGCACGCTGGGACATTAACAAGATCTGTCTCTCAAAAAAAAAAAAAAAAAAAAAA 1819  
QY 1201 GAGAGA 1206  
DB 1820 AAAAAA 1825

RESULT 9  
AC009477/c 173071 bp DNA linear PRI 08-NOV-2000  
LOCUS Homo sapiens BAC clone RP11-209H16 from 2, complete sequence.  
DEFINITION AC009477  
ACCESSION AC009477.4 GI:9755488  
VERSION  
KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 173071)  
Sulston,J.E. and Waterston,R.  
Toward a complete human genome sequence  
Genome Res. 8 (11), 1097-1108 (1998)  
MEDLINE 99063792  
PUBMED 9847074  
REFERENCE 2 (bases 1 to 173071)  
Ali,V., Maupin,R. and Doucette,J.  
The sequence of Homo sapiens BAC clone RP11-209H16  
Unpublished  
REFERENCE 3 (bases 1 to 173071)  
Waterston,R.H.  
Title  
Direct Submission  
Submitted (24-AUG-1999) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
REFERENCE 4 (bases 1 to 173071)  
Waterston,R.H.  
Title  
Direct Submission  
Submitted (09-AUG-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
REFERENCE 5 (bases 1 to 173071)  
Waterston,R.  
Title  
Direct Submission  
Submitted (08-NOV-2000) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Aug 9, 2000 this sequence version replaced gi:7630891.  
COMMENT  
Genome Center  
Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu/gsc  
Contact: saplens@wustl.wustl.edu  
----- Summary Statistics  
Center project name: H\_NH0209H16

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.



**MAPPING INFORMATION:** Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

The RPCR1 human BAC library was made from the blood of one male donor, as described by Ohsagawa, K., Moon, P. Y., Zhao, B., Frøngren, E., Tateno, M., Catanesi, J. J. and de Jong, P. J. (1998). An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>).

Actual start of this clone is at base position 1 of RP11-209H16; actual end is at base position 173071 of RP11-209H16.

## Location/Qualifiers

```

/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-209H16"
/clone_1fb="RPC1-11"
17, 129
repeat_region
/rpt_family="L1"
194, 338
repeat_region
/rpt_family="L1"
343, 662
repeat_region
/rpt_family="Alu"
707, 1014
repeat_region
/rpt_family="L1"
1133, 1499
repeat_region
/rpt_family="MALR"
1761, 1825
repeat_region
/rpt_family="L2"
1965, 2144
repeat_region
/rpt_family="MALR"
2148, 2578
repeat_region
/rpt_family="L1"
2615, 2879
repeat_region
/rpt_family="L1"
2987, 3159
repeat_region
/rpt_family="MER1_type"
5170, 5465
repeat_region
/rpt_family="Alu"
5637, 5943
repeat_region
/rpt_family="Alu"
8914, 8969
repeat_region
/rpt_family="Alu"
8933, 9240
repeat_region
/rpt_family="Alu"
10591, 10700
repeat_region
/rpt_family="MIR"
10735, 11707
repeat_region
/rpt_family="CPG_Island (%GC=74.8, o/e=0.80, #CPGs=113)"
11160, 11360
repeat_region
/rpt_family="Alu"
11355, 11379
repeat_region
/rpt_family="Alu"
11379, 11783
repeat_region
/rpt_family="Alu"
11379, 11783

```

[illegible]



```

misc_feature      21460..21500
                    /note="similar to EST AW953277 (NTD:G8142960)"
misc_feature      21492..21537

Query Match
Best Local Similarity 78.8%; Pred. No. 2,4e-49;
Matches 379; Conservative 0; Mismatches 100; Indels 2; Gaps 1;

14.1%; Score 307; DB 9; Length 173071;

1048 CAGATGATGAAAGCTGCGAAGCGTGGCAGACCGCTATAGTTCTAGCTGTTGGAGGTTG 1107
136573 CAAAAAATGTAAGTTGATGTGTGACATACCTGATCTCAGTACTCGAGGCGCTG 136514
1108 AGGAGGAGGATGCTTGAACAAGGTTTCAAGCCAGCTGGGCAATACAAATC 1167
136513 AGCAGAGAGATGCTTGAAGCCAGAGTTTGAACAGCAGCTGGGCAATATATAGACC 136454
1168 CTGTCTCTCAAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 1227
136453 CTATCTCTAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA 136394
1228 GCCTGATATCCAGCACTTGGGAGGCGAGCGCGGCGGATCATCCTGTGTGAGAGTTT 1287
136393 GCCTGATATCCAGCACTTGGGAGGCGGAGCGGCGGATCATCCTGTGTGAGAGTTT 136334
1288 GAGACCAAGCTGGCAACATGAGCAAAACCCGCTGTGTACTCAAAATGCAAAATTAGCA 1347
136333 AAGACCAAGCTGGCAACATGAGCAAAACCCGCTGTGTACTCAAAATGCAAAATTAGCA 136276
1348 GGGGTGTGTGAGGAGCACTGTGTATCCAGCTACTTGGGAGGCTGAGGAGCAATTCCT 1407
136275 GGGGTGTGTGTGAGGAGCTGTGTATCCAGCTACTTGGGAGGCTGAGGAGCAATTCCT 136216
1408 TGAACCCAGAGAGTGGAGGTTGAGTGAAGTGAAGTGGGCGGTCATCCAGGCTGG 1467
136215 TGAACCTGAGAGTGGAGGTTGAGTGAAGTGAAGTGGGCGGTCATCCAGGCTGG 136156
1468 CGACAGAGCAAGACTCTGTCTCAGAAAAAAGAAAAAAGAAAAAAGAGAGAG 1527
136155 CAACAGAGATGAACTCCGTCTCAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAGAG 136096
1528 A 1528
136095 A 136095

RESULT 10
HS108F13 123101 bp DNA linear PRI 21-MAR-2003
LOCUS Homo sapiens chromosome X BAC CEPH108F13 map Xq13 markers
DEFINITION PHK1-DXS227, complete sequence.
ACCESSION BX295542 AJ239320 AJ239324
VERSION BX295542.1 GI:29150139
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
1 Ramsey,J., Klages,S., Hennig,S., Klein,M., Sudbrak,R., McDonnell,N.,
Chelly,J., Francis,F., Buckle,V., Reinhardt,R. and Lehrach,H.
Unpublished
2 (bases 1 to 123101)
MPING.
Direct Submission
Submitted (15-FEB-1999) MPING, Abt. Lehrach, Max Planck Institut
Fuer Molekulare Genetik, Innesstrasse 73, Berlin, 14195 Germany
config 01
1..123101.
Location/Qualifiers
1..123101
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="X"

```

```

/map="Xq13"
/clone="BAC CEPH108F13"
/clone_lib="CEPHB: human BAC library, creator: Denis Le
Paslier"
/note="region between markers PHK1-DXS227"

ORIGIN

Query Match
Best Local Similarity 78.1%; Pred. No. 4,5e-49;
Matches 381; Conservative 0; Mismatches 104; Indels 3; Gaps 1;

14.1%; Score 305.6; DB 9; Length 123101;

1021 TAACTATATAGAGATTGTGACAAAACAGATGATTAAGTCCGAACCTGGCAACGC 1080
48208 TGAATATATATATATAGTGTGACATATATATATATATATATATATATATATATAT 48267
1081 TCATAGTTTATAGTGTGAGGAGTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1140
48268 CTATATCTTACACCTCTGGAGAGCTGAGGTGGAGAGATCCTGAGGCGGAGTTTCA 48327
1141 GGCAGCCTGGCAACATTAACAGATCTGTCTCTCAAAAAAAGAAAAAAGAAAAAAGAAAA 1200
48328 GACCAAGCTGGCAACATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 48387
1201 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
48388 CAGAGTGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 48447
1261 GGGCGGATACCTGTGTGTGAGAGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
48448 AGGTAGGTACCTGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 48507
1321 CTGTACTCAAAATGCAAAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
48508 CTGTACTCAAAATGCAAAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 48567
1381 TTGGAGAGCTGA--GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1437
48568 TCGGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 48627
1438 TGAAGATGTCCTGTCATCCAGCTGGGCGCAAGAGCAAGAGAGAGAGAGAGAGAGAG 1497
48628 TGAAGATGTCATTCATCTCCAGCTGGGCGCAAGAGATGAAGATCCGTTTCAAAAAA 48687
1498 AAAAAAA 1505
48688 AAAAAAA 48695

RESULT 11
AC084024/c 135404 bp DNA linear PRI 19-SEP-2002
LOCUS Homo sapiens chromosome 8, clone RP11-90P5, complete sequence.
DEFINITION AC084024
ACCESSION AC084024
VERSION AC084024.17 GI:23196607
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
1 Birren,B., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 8, clone RP11-90P5
Unpublished
2 (bases 1 to 135404)
Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavsky,L.,
Boukhalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collins,A., Cooke,P.,
DeArliano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
Graham,L., Grand-Pierre,N., Hagos,B., Hesford,A., Horton,L.,
Ilier,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRoque,K.,
Lamazaree,R., Lander,T., Lemoczky,J., Levine,R., Liu,C., Liu,G.,

```



Macdonald, P., Margulis, N., McCarthy, M., McEwan, P., McKernan, K., McSheehy, R., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Rieback, M., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Sougniez, C., Spencer, B., Strange-Thomann, N., Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

Submitted (10-OCT-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 135404)  
Barren, B., Nuebaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Baran, N., Bastien, V., Bloom, T., Boguslavsky, L., Bouknight, B., Camarata, U., Chang, J., Chararo, B., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fero, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardina, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kanat, A., Kartas, A., Kells, C., Landers, T., Levine, R., Lindblad-Ton, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Strange-Thomann, N., Stojanovic, N., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zemek, L., Zimmer, A. and Zody, M.

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

Submitted (06-JUL-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
4 (bases 1 to 135404)  
Barren, B., Nuebaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Baran, N., Bastien, V., Bloom, T., Boguslavsky, L., Bouknight, B., Camarata, U., Chang, J., Chararo, B., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fero, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardina, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kanat, A., Kartas, A., Kells, C., Landers, T., Levine, R., Lindblad-Ton, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Strange-Thomann, N., Stojanovic, N., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zemek, L., Zimmer, A. and Zody, M.

## TITLE

## JOURNAL

## COMMENT

Submitted (19-SEP-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Sep 19, 2002 this sequence version replaced gi:21700615.  
Clone is not repeat masked.

## Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

## Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence.submissions@genome.wi.mit.edu](mailto:sequence.submissions@genome.wi.mit.edu)

## Project Information

Center project name: L11299

Center clone name: 90\_P\_5

## FEATURES

## source

1. 135404  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"

/db\_xref="taxon:9606"  
/chromosome="8"  
/map="8"  
/clone="RP11-90P5"  
/clone\_lib="RP11-11 Human Male BAC"  
41330..41334  
/note="30 qual SINGL region"  
41355..41359  
/note="30 qual SINGL region"  
41400..41405  
/note="30 qual SINGL region"  
5396..5406  
/note="30 qual SINGL region"  
54023..54027  
/note="30 qual SINGL region"  
54033..54047  
/note="single clone coverage"  
60716..60745  
/note="30 qual SINGL region"  
60755..60766  
/note="30 qual SINGL region"  
99957..99969  
/note="Probably 13 T s. possibly 11 or 12."  
134341  
/note="probably T, possibly C"

## ORIGIN

## Query Match

Best Local Similarity 14.0%; Score 304.6; DB 9; Length 135404;  
Matches 361; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

1058 AACCTCGAACCGTGGACACGCTCATAGTTTACTGCTTGGAGTGGAGGAGG 1117  
93059 AGCCGAGCAGCAGTGGCTCATGCTGTATCCACGACTTGGAGCCCAAGTGGAGA 93000  
1118 ATGGCTTGAACAAGGTGTTCAAGCCAGCCTGGCAATTAAGATCTGTCTTCA 1177  
92999 ATGATTGAGTCAAGAGTTTGAACGACGCTGACACATATGAGACCTGTTTTCAC 92940  
1178 AAAAAAAAAAAAAAAAAAAGAGAGAGGCGCGGCTGTGTCTCAAGCTTATTC 1237  
92939 AAAAAAAAAAATAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTA 92880  
1238 CCAAGCACTTGGAGGCGAGCGAGCGAGATCACTGTTGTCAGAGTTTGAACCAAGC 1297  
92879 CTGCACTTGGAGGCGAGCGAGCGAGTGTGCTGAGGTGAGATTTTGAACCAAGC 92820  
1298 TGCCCAACATGGCAAAACCCGCTGTACTCAAAATGCAAAATTAAGCCAGGCGTGTAG 1357  
92819 TGCCCAAAATGGGAACCCGATCTCTAATAAATAAATAAATAAATAAATAAATAA 92760  
1358 CAGGCACTGTATCCAGCTACTTGGAGGCTGAGGCGAGGAGATCGTTAACCAGAG 1417  
92759 CAGGCGCTGTATCCAGCTACTCGGCGGCTGAGGAGGAGATCGTTAACCAGAG 92700  
1418 AGGTGAGGTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1477  
92699 AGCGGAGGTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 92640  
1478 AAGACTGTGTCAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG 1512  
92639 AAGACTGTGTCAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG 92605

RESULT 12  
AC024242/c  
LOCUS  
DEFINITION  
Homo sapiens chromosome 8 clone RP11-513b5, WORKING DRAFT SEQUENCE,  
17 unordered pieces.  
AC024242  
AC024242.3 GI:8570266  
VERSION  
HTG; HTGS PHASE1; HTGS\_DRAFT.  
KEYWORDS  
Homo sapiens (human)







```
source
1. .195623
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="8"
/map="8q23"
/clone="RB1198A1"
/cell_line="FLEB 14 - 14"
/cell_type="pre-pro-B cell"
/clone_lib="Keio BAC library"
repeat_region
376. .429
/evidence=not_experimental
/rpt_family="GA-rich"
1725. .1748
/evidence=not_experimental
/rpt_family="AT-rich"
1749. .2040
/evidence=not_experimental
/rpt_family="AluSx"
2041. .2071
/evidence=not_experimental
/rpt_family="AT-rich"
2261. .2322
/evidence=not_experimental
/rpt_family="L1MC5"
complement(2323. .2613)
/evidence=not_experimental
/rpt_family="AluDo"
2614. .2740
/evidence=not_experimental
/rpt_family="L1MC5"
2741. .3036
/evidence=not_experimental
/rpt_family="AluSx"
3040. .3330
/evidence=not_experimental
/rpt_family="AluY"
3331. .3350
/evidence=not_experimental
/rpt_family="L1MC5"
3351. .3653
/evidence=not_experimental
/rpt_family="AluDb"
3654. .3738
/evidence=not_experimental
/rpt_family="L1MC5"
complement(3739. .4244)
/evidence=not_experimental
/rpt_family="MLT1D"
4245. .4630
/evidence=not_experimental
/rpt_family="AluSp"
4631. .4716
/evidence=not_experimental
/rpt_family="AT-rich"
complement(4716. .5549)
/evidence=not_experimental
/rpt_family="L1MC4"
5549. .5622
/evidence=not_experimental
/rpt_family="TTRTA)n"
complement(5625. .5713)
/evidence=not_experimental
/rpt_family="L1MC4"
5942. .5996
/evidence=not_experimental
/rpt_family="AT-rich"
6056. .6339
/evidence=not_experimental
/rpt_family="AluSx"
6340. .6375
/evidence=not_experimental
/rpt_family="TAA)n"

repeat_region
6395. .6448
/evidence=not_experimental
/rpt_family="Alu"
6449. .6759
/evidence=not_experimental
/rpt_family="L1MB7"
6760. .7085
/evidence=not_experimental
/rpt_family="AluDo"
7086. .7242
/evidence=not_experimental
/rpt_family="L1MB7"
7243. .7707
/evidence=not_experimental
/rpt_family="MLT1D"
7708. .7736
/evidence=not_experimental
/rpt_family="L1MB7"
7741. .7862
/evidence=not_experimental
/rpt_family="L1MB7"
complement(7863. .8276)
/evidence=not_experimental
/rpt_family="MSTB"
8277. .8294
/evidence=not_experimental
/rpt_family="L1MB7"
8473. .9355
/evidence=not_experimental
/rpt_family="L1MB5"
complement(9370. .9867)
/evidence=not_experimental
/rpt_family="AluSx"
complement(10103. .10388)
/evidence=not_experimental
/rpt_family="AluDb"
10799. .11558
/evidence=not_experimental
/rpt_family="L1R39"
11590. .11704
/evidence=not_experimental
/rpt_family="AluDo"
11705. .11998
/evidence=not_experimental
/rpt_family="AluY"
11999. .12139
/evidence=not_experimental
/rpt_family="AluDo"
12173. .12248
/evidence=not_experimental
/rpt_family="AluDo/FLAM"
12262. .12299
/evidence=not_experimental
/rpt_family="AT-rich"
12371. .12680
/evidence=not_experimental
/rpt_family="TMB1B"
12808. .12949
/evidence=not_experimental
/rpt_family="I2"
13337. .13476
/evidence=not_experimental
/rpt_family="AluSx"
13477. .13768
/evidence=not_experimental
/rpt_family="AluSp"
13769. .13944
/evidence=not_experimental
/rpt_family="AluSx"
14073. .14197
/evidence=not_experimental
/rpt_family="I2"
14594. .14731
```



```

/evidence=not_experimental
/rpt_family="LIME1"
repeat_region 14961..15034 /evidence=not_experimental
/rpt_family="MER104"
repeat_region 15035..15076 /evidence=not_experimental
/rpt_family="AT_rich"
repeat_region 15077..15196 /evidence=not_experimental
/rpt_family="MER104"
repeat_region 15853..15941 /evidence=not_experimental
/rpt_family="MIR"
repeat_region 16086..16181 /evidence=not_experimental
/rpt_family="MIR"
repeat_region 16306..16328 /evidence=not_experimental
/rpt_family="MIR"
repeat_region 16493..16719 /evidence=not_experimental
/rpt_family="AT_rich"
repeat_region 16680..16729 /evidence=not_experimental
/rpt_family="MIR"
repeat_region 17013..17185 /evidence=not_experimental
/rpt_family="L2"
repeat_region 17186..17530 /evidence=not_experimental
/rpt_family="MER47A"
repeat_region 17186..17530 /evidence=not_experimental
/rpt_family="L1PA16"

Query Match 14.0%; Score 304.6; DB 9; Length 195623;
Best Local Similarity 75.3%; Pred. No. 6.8e-49;
Matches 379; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 1039 GTGACAAAACAGATGATTAAGCTGCGACCGTGCACACCTCATAGTTCTACTGCTT 1098
Db 73446 GTTAAAAATCACTATATTATGGCTGGCTGGCTGCTATGCTGTAATCTTACACTTT 73387
QY 1099 GCGAGTTGAGGAGGAGGAGATGCTTGAACACAGGTGTTCAAGCCGACCTGGCAACAT 1158
Db 73386 GGAAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 73327
QY 1159 AACAAAGATCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1218
Db 73326 AGGAGACCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 73267
QY 1219 GTGGCTACGCTCTGTAATCCAGCACCTTTGGAGGCGGAGCGGCGGATCACTGTGCT 1278
Db 73266 GTGGCTACGCTCTGTAATCCAGCACCTTTGGAGGCGGAGCGGCGGATCACTGTGCT 73207
QY 1279 CAGAGTTTGAACACAGCTGCGCAACATGCGCAAAACCCGCTGTAATCTCAAAATGAAA 1338
Db 73206 CAGAGTTTGAACACAGCTGCGCAACATGCGCAAAACCCGCTGTAATCTCAAAATGAAA 73147
QY 1339 AATTAGCAGGCGGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1398
Db 73146 AATTAGCAGGCGGTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 73087
QY 1399 AGAATCGCTTGAACCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1458
Db 73086 AGAATCGCTTGAACCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 73027
QY 1459 CAGCTTGGCGGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1518
Db 73026 CAGCTTGGCGGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 72967
QY 1519 GAGAAAGAGCAATATTTGGGA 1541
Db 72966 ATAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 72944

```

```

RESULT 14
AC091164/c
LOCUS
DEFINITION
AC091164 bp DNA linear HTG 03-APR-2001
SEQUENCE, 6 unordered pieces.
AC091164.1 GI:13518148
VERSION
AC091164.1 HTG; HTGS PHASE1; HTGS_DRAFT.
KEYWORDS
Homo sapiens (human)
SOURCE
Homo sapiens
Organism
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 172884)
2 (bases 1 to 172884)
REFERENCE
AUTHORS
Bairren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE
Homo sapiens chromosome 8, clone RP11-266D22
JOURNAL
REFERENCE
AUTHORS
2 (bases 1 to 172884)
Bairren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Boguslavsky,L., Boukhalter,B., Brown,A.,
Camarata,J., Campoliano,A., Chang,J., Choepel,Y., Colangelo,M.,
Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K.,
Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Hagos,B., Harford,A., Horton,L., Hulme,W.,
Iliev,I., Johnson,R., Jones,C., Karatas,A., Labocque,K.,
Lamarez,R., Landers,T., Lehotzky,J., Levine,R., Liu,G.,
Maclean,C., MacDonald,P., Margulis,N., Matthews,C., McCarthy,M.,
McEwan,P., McKernan,K., McPherson,R., Melidim,J., Menus,L.,
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
Retter,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schnepf,R., Seaman,S.,
Severy,P., Sounguez,C., Spencer,B., Strange-Thomann,N.,
Stojanovic,N., Straus,N., Subramanian,A., Talamas,J., Testave,S.,
Theodore,J., Travers,M., Travis,N., Trigglio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,X.
Direct Submission
Submitted (03-APR-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L11677
Center clone name: 266.D.22
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly Program: Phrap; version 0.960731
Consensus quality: 171435 bases at least Q40
Consensus quality: 172085 bases at least Q30
Consensus quality: 172251 bases at least Q20
Insert size: 176000; agarose-fp
Insert size: 172384; sum-of-contigs
Quality coverage: 8.9 in Q20 bases; agarose-fp
Quality coverage: 9.1 in Q20 bases; sum-of-contigs

```

NOTE: This is a 'working draft' sequence. It currently consists of 6 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.



```

*      1      1787: contig of 1787 bp in length
*      1788      1887: gap of 100 bp
*      1888      15117: contig of 13230 bp in length
*      15118      15217: gap of 100 bp
*      15218      37920: contig of 22703 bp in length
*      37921      38020: gap of 100 bp
*      38021      79049: contig of 41029 bp in length
*      79050      79149: gap of 100 bp
*      79150      131081: contig of 51932 bp in length
*      131082      131181: gap of 100 bp
*      131182      172884: contig of 41703 bp in length.
FEATURES
SOURCE
1..172884
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="RP11-266D22"
/clone_1b="RPC1-11 Human Male BAC"
misc_feature
1..1787
/clone="assembly_fragment"
vector_side:left"
misc_feature
1888..15117
/clone="assembly_fragment"
misc_feature
15218..37920
/clone="assembly_fragment"
misc_feature
38021..79049
/clone="assembly_fragment"
misc_feature
79150..131081
/clone="assembly_fragment"
misc_feature
131182..172884
/clone="assembly_fragment"
vector_side:right"
ORIGIN
Query Match      14.0%; Score 303.6; DB 2; Length 172884;
Best Local Similarity 77.1%; Pred. No. 11e-48;
Matches 387; Conservative 0; Mismatches 104; Indels 11; Gaps 1;
QY      1060 GCTGGCAACCGTGGACACGCTCATAGTCTTCTGAGTGGAGGTTGAGAGGAGGAT 1119
      |||||
DB      60049 GCTGGCGCCAGTGGTCACTGTAATCTGACGACTTTGGAGGCTGAGTGGAGGAT 59990
QY      1120 GCGTTGAACACAGTGTTCAGGCCAGCGCTGGGCAATACAAAGATCTGTCTCTCAA 1179
      |||||
DB      59989 CACTTGAGCTCAGAGTTCGAGACAGCGCTGGGCAATAGAGTGAAGCCCATCTCTATT 59930
QY      1180 AAAAAAAAA-----AAAAAAAAAGAGAGGCGCGGCGCTGTGCTCAGC 1228
      |||||
DB      59929 TTTAATAAATAATAATATTTTAAAAAATAAATAAAGGCGCAGTCCGTGCTCAAG 59870
QY      1229 CCTGTATCCAGACCTTTGGAGGCGGAGCGGCGGATCACTGTGTGAGAGTTG 1288
      |||||
DB      59869 CCTGTATCCAGACCTTTGGAGGCGGAGCGGCGGATCACTGTGTGAGAGTTG 59810
QY      1289 AGACCAAGCTGGCCCAACATGCAAAACCCCGTCTGTACTCAAAATGCAAAATTAACCAAG 1348
      |||||
DB      59809 AATACCAAGCTGGCCCAACATGTAACCCCGTCTGTACTCAAAATTAATTAAGTCAG 59750
QY      1349 GCGTGTAGCAGGACCTGTATCCAGTCTTCTGGAGGCTGAGGCGAGGATCGTT 1408
      |||||
DB      59749 GTGTGTGTCAGGCGCGCTGTATCCAGTCTCTGCGAGGCTGAGGCGAGGAATCGCTT 59690
QY      1409 GAACCCAGAGGTGAGAGTTCAGATGAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGC 1468
      |||||
DB      59689 GAACGTGGAGGTGAGAGTTCAGATGAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGC 59630
QY      1469 GACAAGAGCAAGACTGTCTCAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1528
      |||||
DB      59629 GACAAGAGCAAGACTGTCTCAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 59570

```

```

QY      1529 ACAATATTTGGAGAGAGAGAT 1550
      |||||
DB      59569 TTAAGTTCTCCAGATAAGAT 59548
RESULT 15
AC024035
LOCUS
DEFINITION
ACCESSION
AC024035
VERSION
AC024035.3 GI:7637351
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 176063)
Waterston,R.H.
REFERENCE
1 (bases 1 to 176063)
Waterston,R.H.
AUTHORS
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 176063)
Waterston,R.H.
DIRECT SUBMISSION
Submitted (20-FEB-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Apr 22, 2000 this sequence version replaced gi:7024222.
COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Project Information
Center project name: H_NH0266D22
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing method: Dye-terminator Big Dye; 0% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.950319
Consensus quality: 170547 bases at least Q40
Consensus quality: 171287 bases at least Q30
Consensus quality: 171861 bases at least Q20
Insert size: 151000; agarose-fp
Insert size: 175063; sum-of-coverage
Quality coverage: 6.16 in Q20 bases; sum-of-coverage
Quality coverage: 6.14 in Q20 bases; sum-of-coverage
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
2493: contig of 2493 bp in length
2494      2593: gap of unknown length
2594      7948: contig of 5355 bp in length
7949      8048: gap of unknown length
8049      13046: contig of 4998 bp in length
13047      13146: gap of unknown length
13147      19902: contig of 6756 bp in length
19903      20002: gap of unknown length
20003      27779: contig of 7777 bp in length
27780      27879: gap of unknown length
27880      36679: gap of 8800 bp in length
36680      36779: gap of unknown length
36780      52036: contig of 15557 bp in length
52037      52136: gap of unknown length
52137      71975: contig of 19839 bp in length

```







GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

CM nucleic - nucleic search, using sw model

Run on: March 12, 2004, 12:26:39 ; Search time 869.74 Seconds

(without alignments)  
10609.014 Million cell updates/sec

Title: US-09-966-880a-15

Perfect score: 2172  
Sequence: 1 cccctgatacggtgatga.....agatgatacgataaactgt 2172

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database: N\_Geneseq\_29Jan04:\*

1: geneseqn1980s:\*\n2: geneseqn1990s:\*\n3: geneseqn2000s:\*\n4: geneseqn2001s:\*\n5: geneseqn2001bs:\*\n6: geneseqn2002s:\*\n7: geneseqn2003as:\*\n8: geneseqn2003bs:\*\n9: geneseqn2003cs:\*\n10: geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2172	100.0	2172	3	AAC55319 Human act
2	2172	100.0	2791	6	AB573287 DNA encod
3	2172	100.0	2791	6	AB573288 DNA encod
4	2172	100.0	2818	3	AAC55312 Human act
5	2172	100.0	6564	3	AAC55314 Human act
6	2172	100.0	11204	3	AAC55339 Human act
7	2172	100.0	11204	6	AB573286 DNA encod
8	1488	68.5	1543	7	ABX81088 Human imm
9	923	42.5	1543	7	ABX05468 Human nov
10	574	26.4	574	4	AAK81089 Human imm
11	429	19.8	429	4	AAK61819 Human imm
12	295.2	13.6	13862	4	AAI02789 Human rep
13	295.2	13.6	13862	4	AAI07516 Human rep
14	295.2	13.6	13862	4	ABX08208 Human cva
15	293.8	13.5	110000	7	ACP42745_0 Human ALM
16	292.6	13.5	7809	5	AAI03406 Human rep
17	292.6	13.5	7809	5	ABAI15724 Human ner
18	292.4	13.5	30393	4	AAK67239 Human imm
19	291	13.4	25012	5	ABAI15431 Human ner
20	291	13.4	25012	7	AB573798 Secreted
21	291	13.4	25012	7	AD598470 Human sec
22	288.8	13.3	96596	8	ADA02864 Human TTP
23	288.8	13.3	96596	9	ADB72602 Human ITP

C	24	288.8	13.3	96596	9	ADC85343	ADC85343	Mouse ITP
C	25	287.8	13.3	110000	9	AD511169_0	AD511169_0	Human tra
C	26	286.2	13.2	37314	6	AAK71358	AAK71358	Human tra
C	27	286.2	13.2	172570	6	ABQ88207	ABQ88207	Human rep
C	28	284.4	13.1	32249	4	AAI04676	AAI04676	Human rep
C	29	284.4	13.1	32249	4	AB197583	AB197583	Human tes
C	30	284.4	13.1	129722	6	ABQ88117	ABQ88117	Human ost
C	31	283.4	13.0	1792	6	ABA96614	ABA96614	Human alp
C	32	283.2	13.0	22428	7	AA541759	AA541759	Genomic s
C	33	283.2	13.0	22428	7	AB267767	AB267767	Human sec
C	34	283.2	13.0	22428	7	AB274201	AB274201	Secreted
C	35	283.2	13.0	22428	7	ADA98730	ADA98730	Human sec
C	36	282.4	13.0	13919	6	ABK86218	ABK86218	DNA encod
C	37	282.4	13.0	13919	6	ABK86221	ABK86221	Alp-1/FLA
C	38	282.4	13.0	13919	6	ABK86220	ABK86220	Alp-1/FLA
C	39	282.4	13.0	21404	6	ABK86229	ABK86229	Alp-1/FLA
C	40	281.4	13.0	65608	6	AB162910	AB162910	Breast ca
C	41	281.4	13.0	65608	6	AB164414	AB164414	Stomach C
C	42	281.4	13.0	65608	6	AB167668	AB167668	Oesophagu
C	43	281	12.9	23452	4	AA542122	AA542122	Genomic s
C	44	280.2	12.9	11821	4	AA136492	AA136492	Human mus
C	45	280.2	12.9	11821	7	ABX59480	ABX59480	CDNA enco

#### ALIGNMENTS

RESULT 1	AAC55319	AAC55319 standard; DNA; 2172 BP.
ID	AAC55319;	
AC	AAC55319;	
XX		
DT	05-FEB-2001 (first entry)	
XX		
DE	Human activation-induced cytidine deaminase exon 5 SEQ ID NO:15.	
XX		
KW	Activation-induced cytidine deaminase; AID; cytidine deaminase; immune related disease; allergy; allergic disease; antiallergic; antidiabetic; antidiabetic; ophthalmological; anti-HIV; dermatological; gene therapy; B cell associated immune system disorder; food allergy; immunodeficiency disease; immunoglobulin A deficiency disease; asthma; IGA nephritis; gamma-globulinemia; atopic dermatitis; allergic colitis; drug allergy; allergic rhinitis; Rosen disease; Digeorge disease; AIDS; ataxia telangiectasia; common variable immunodeficiency disorder; major histocompatibility class II deficiency disease; auto immunodeficiency syndrome; Igg subclass selection disorder; ds.	
KW		
XX		
OS	Homo sapiens.	
XX		
PN	W0200058480-A1.	
XX		
PD	05-OCT-2000.	
XX		
PF	28-MAR-2000; 2000MO-JP001918.	
XX		
PR	29-MAR-1999; 990P-00087192.	
PR	24-JUN-1999; 990P-00178999.	
PR	27-DEC-1999; 990P-00371382.	
XX		
PA	(NISR) JAPAN TOBACCO INC.	
PA	(HONU) HONTO T.	
PI	Honjo T, Muramatsu M;	
XX		
DR	WPI; 2000-611715/58.	
XX		
PT	Nucleic acid encoding activation induced cytidine deaminase, useful as a target for drug development for immune-related diseases including allergies.	
PT		
XX		
PS	Claim 16; Page 152-153; 174pp; Japanese.	



CC The present invention describes an activation-induced cytidine deaminase  
CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and has  
CC cytidine activity similar to APOBEC-1. AID has antiallergic, antianemic,  
CC antiasthmatic, ophthalmological, anti-HIV and dermatological activities,  
CC and can be used in gene therapy. AID polynucleotides are useful in  
CC methods for identifying drugs for the treatment of B cell associated  
CC immune system disorders, immunodeficiency diseases and allergies, such as  
CC immunoglobulin A (IgA) deficiency disease, IgA nephritis, gamma-  
CC globulinemia, atopic dermatitis, allergic colitis, asthma, food allergy,  
CC drug allergy, allergic rhinitis, Rosen disease, Digeorge disease, ataxia  
CC telangiectasia, common variable immunodeficiency disorder, MHC (major  
CC histocompatibility class) class II deficiency disease, AIDS (auto-  
CC immunodeficiency syndrome), elevated IgE disorder, and IgG subclasses  
CC selection disorder. The DNA sequences encoding AID may be used for gene  
CC therapy and the antibodies to the AID protein may be used for diagnosis  
CC and treatment of these disorders. The present sequence represents the  
CC exon 5 genomic DNA sequence of human AID  
XX

SQ Sequence 2172 BP; 702 A; 379 C; 465 G; 626 T; 0 U; 0 Other;

Query Match 100.0%; Score 2172; DB 3; Length 2172;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCCTGATAGAGTGAATGACTTACGAGACGCAATTCGACTTTGGACCTTGTATGACAA 60  
DB 1 CCCCTGATAGAGTGAATGACTTACGAGACGCAATTCGACTTTGGACCTTGTATGACAA 60

QY 61 CTTCCAGGAATGTCACACAGTGAATATCTCTGCTGAAGACAGTGAATAAACAAGT 120  
DB 61 CTTCCAGGAATGTCACACAGTGAATATCTCTGCTGAAGACAGTGAATAAACAAGT 120

QY 121 CCTTCAAGCTCTCTGTTTATCTTCAACTCTCACTCTTCTAGAGTTAGAGAAAA 180  
DB 121 CCTTCAAGCTCTCTGTTTATCTTCAACTCTCACTCTTCTAGAGTTAGAGAAAA 180

QY 181 AATTTATATAGACTCTTTTAAAAAGATCTATGCTTGAATAATGAGAGGACACAGT 240  
DB 181 AATTTATATAGACTCTTTTAAAAAGATCTATGCTTGAATAATGAGAGGACACAGT 240

QY 241 CTGGCCAGGAGCTGCTGCAATGCTGCAATTTGAATGCAACATGTCCTTCTGGA 300  
DB 241 CTGGCCAGGAGCTGCTGCAATGCTGCAATTTGAATGCAACATGTCCTTCTGGA 300

QY 301 AATAAGAGACTGAGAGACTGAGAGCACTAAAGTCAACGTTTTCTATGACTTTTA 360  
DB 301 AATAAGAGACTGAGAGACTGAGAGCACTAAAGTCAACGTTTTCTATGACTTTTA 360

QY 361 GGTAGGATGAGAGAGAGTATCTTAAAAAGCATGGTGAAGATCAATGTTTTTA 420  
DB 361 GGTAGGATGAGAGAGAGTATCTTAAAAAGCATGGTGAAGATCAATGTTTTTA 420

QY 421 TATCAACATCCTTTATTTATTTGATTCATTTGATTAACAGTGGTGTATGATAGATTT 480  
DB 421 TATCAACATCCTTTATTTATTTGATTCATTTGATTAACAGTGGTGTATGATAGATTT 480

QY 481 TCTATTTCTTTTCCCTTGAAGTCTTCTTCAAGTAAACAACCTCTCATAGGCCATGA 540  
DB 481 TCTATTTCTTTTCCCTTGAAGTCTTCTTCAAGTAAACAACCTCTCATAGGCCATGA 540

QY 541 TCTATATGAGCTCTTAAAGAGTATCTGGGATTTGACCCCAACCATCTCTCCAAA 600  
DB 541 TCTATATGAGCTCTTAAAGAGTATCTGGGATTTGACCCCAACCATCTCTCCAAA 600

QY 601 GCATTATATCCATATGCGCTGTATGTTTAAATCAGACAGAGCATGTTTTATGTTTG 660  
DB 601 GCATTATATCCATATGCGCTGTATGTTTAAATCAGACAGAGCATGTTTTATGTTTG 660

QY 661 TACAAAAGAGATGTTATGGGTGGGATGAGGATATGACCATGCGTACCTTCAA 720  
DB 661 TACAAAAGAGATGTTATGGGTGGGATGAGGATATGACCATGCGTACCTTCAA 720

QY 721 GGTACTTATATAAGATCTTAAATGGGACAGAGAGCTGTGAACAACACCCATAATA 780

DB 721 GGTACTTATATAAGATCTTAAATGGGACAGAGAGCTGTGAACAACACCCATAATA 780

QY 781 TGGGTGATGCTGTAAGTAGCAAAATCTTGTGAAAACGAAACCTTTTAAAGAAATCCCT 840  
DB 781 TGGGTGATGCTGTAAGTAGCAAAATCTTGTGAAAACGAAACCTTTTAAAGAAATCCCT 840

QY 841 AATTTGAAAACCCCAACCTTCAATATCATATTTGCAACCAATTGAAAGAGTTG 900  
DB 841 AATTTGAAAACCCCAACCTTCAATATCATATTTGCAACCAATTGAAAGAGTTG 900

QY 901 CTTGAATGTTGGGAGAGAGAAAATCTATGGCTCGGGGCTCTTCAATCTCAAGATG 960  
DB 901 CTTGAATGTTGGGAGAGAGAAAATCTATGGCTCGGGGCTCTTCAATCTCAAGATG 960

QY 961 CCAATCAGGTCAGAGTTTGTCTCATTTTGTATGTTGATGATCTTCCCAAGTATAT 1020  
DB 961 CCAATCAGGTCAGAGTTTGTCTCATTTTGTATGTTGATGATCTTCCCAAGTATAT 1020

QY 1021 TAACTATATAGAGATGCTGAACAACAAGATTAAGCTGCAACCTGGACACGC 1080  
DB 1021 TAACTATATAGAGATGCTGAACAACAAGATTAAGCTGCAACCTGGACACGC 1080

QY 1081 TCAATGTTCTAGCTGCTTGGAGGTTGAGAGAGAGATGGCTTGAACACAGTGTCAA 1140  
DB 1081 TCAATGTTCTAGCTGCTTGGAGGTTGAGAGAGAGATGGCTTGAACACAGTGTCAA 1140

QY 1141 GGCACAGCTGGGCAACATPACAAAGATCCTGTCTCAAAAAAAGAAAAAGAAA 1200  
DB 1141 GGCACAGCTGGGCAACATPACAAAGATCCTGTCTCAAAAAAAGAAAAAGAAA 1200

QY 1201 GAGAGAGGCGGAGGCTGCTGCTGCAACGCTGTAAATCCAGACCTTTGGAGCGGACCC 1260  
DB 1201 GAGAGAGGCGGAGGCTGCTGCTGCAACGCTGTAAATCCAGACCTTTGGAGCGGACCC 1260

QY 1261 GGGCGGATCAGCTGTGCTGAGAGGTTGAGACCAAGCTGGCCACATGGCAAAACCCGCT 1320  
DB 1261 GGGCGGATCAGCTGTGCTGAGAGGTTGAGACCAAGCTGGCCACATGGCAAAACCCGCT 1320

QY 1321 CTGTACTCAAAATGCAAAAATTAGCAGGCGGTGAGAGAGGACCTGTATGCCAGCTAC 1380  
DB 1321 CTGTACTCAAAATGCAAAAATTAGCAGGCGGTGAGAGAGGACCTGTATGCCAGCTAC 1380

QY 1381 TTGGAGGCTGAGAGAGAGATCCCTTGAACCCAGAGGTGAGAGTTCAGTAAGCTGA 1440  
DB 1381 TTGGAGGCTGAGAGAGAGATCCCTTGAACCCAGAGGTGAGAGTTCAGTAAGCTGA 1440

QY 1441 GATCGTCCGTTGCACTCCAGGCTGGGCGACAAGAGCAAGCTGTCTCAGAAAAAAA 1500  
DB 1441 GATCGTCCGTTGCACTCCAGGCTGGGCGACAAGAGCAAGCTGTCTCAGAAAAAAA 1500

QY 1501 AAAAAAAGAGAGAGAGAGAAAGAAACAATATTTGGAGAGAGAGATGGGGAGCAT 1560  
DB 1501 AAAAAAAGAGAGAGAGAGAAAGAAACAATATTTGGAGAGAGAGATGGGGAGCAT 1560

QY 1561 TGCAGAGAAATGTGCTTATATCCAAATATTAAGAGCCAAATAGGAGATCCCTATTTG 1620  
DB 1561 TGCAGAGAAATGTGCTTATATCCAAATATTAAGAGCCAAATAGGAGATCCCTATTTG 1620

QY 1621 TCTCTTTGGTGTCTATTTTGTCCCTAACACTGTCTTGAACAGTGAAGAAAAATATTCGA 1680  
DB 1621 TCTCTTTGGTGTCTATTTTGTCCCTAACACTGTCTTGAACAGTGAAGAAAAATATTCGA 1680

QY 1681 ATACCATATCCCTGTGCGGTATTAACCTAGCAACCTTGCATAGATGAGCATCC 1740  
DB 1681 ATACCATATCCCTGTGCGGTATTAACCTAGCAACCTTGCATAGATGAGCATCC 1740

QY 1741 ACAGGAAACTGAAAGCAACCTGCTATTTATCTATATGTAATGTAATGTTGTAA 1800  
DB 1741 ACAGGAAACTGAAAGCAACCTGCTATTTATCTATATGTAATGTAATGTTGTAA 1800

QY 1801 AGAGTTAAAAATGTTACTTCAATGATTAATTAATTAATTAATTAATTAATG 1860



DB 1801 AGAGTAAATGTTACTGATGATTCATTATATTTATTTATTTGCGTCTATG 1860  
 QY 1861 ATTTTATTAACATGATTTCTTTCTGATATATTTGAATGAGTCTCAAGCTTATA 1920  
 DB 1861 ATTTTATTAACATGATTTCTTTCTGATATATTTGAATGAGTCTCAAGCTTATA 1920  
 QY 1921 AATTTTAACTTTAGAAATGATTTCTTAATTAACAGATGATGATTTGATGATGAT 1980  
 DB 1921 AATTTTAACTTTAGAAATGATTTCTTAATTAACAGATGATGATTTGATGATGAT 1980  
 QY 1981 GATGCTACGAGCATTCTCTGATTTTGTAGTAACTTTATGACAGCAATTTGCTC 2040  
 DB 1981 GATGCTACGAGCATTCTCTGATTTTGTAGTAACTTTATGACAGCAATTTGCTC 2040  
 QY 2041 TGGCTCACTTTCAATGATGATTAATTAATTAATTTGGAAGCTGTGAGATATAA 2100  
 DB 2041 TGGCTCACTTTCAATGATGATTAATTAATTAATTTGGAAGCTGTGAGATATAA 2100  
 QY 2101 TACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2160  
 DB 2101 TACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2160  
 QY 2161 GGAATTAATTTG 2172  
 DB 2161 GGAATTAATTTG 2172  
 RESULT 2  
 ABS73287  
 ID ABS73287 standard, DNA; 2791 BP.  
 XX  
 AC ABS73287;  
 XX  
 DT 04-DEC-2002 (first entry)  
 XX  
 DE DNA encoding human translocation del(12p) protein #2.  
 XX  
 KM Chromosome aberration; oncogenic fusion protein; cancer;  
 KM proliferative disease; cellular fusion isoform; heat shock protein 90;  
 KM HSP-90; rheumatoid arthritis; cancer; haematopoietic disorder;  
 KM T cell lymphoma; B cell lymphoma; chronic myeloid leukaemia; CML;  
 KM acute myeloid leukaemia; AML; chronic myelomonocytic leukaemia; CMML;  
 KM acute lymphoblastic leukaemia; ALL; APL; NHL; solid tumour;  
 KM papillary thyroid carcinoma; Ewing's sarcoma; melanoma; liposarcoma;  
 KM rhabdomyosarcoma; synovial sarcoma; viral infection; gene; ds.  
 OS Homo sapiens.  
 PN MO20026900-A2.  
 PD 12-SEP-2002.  
 PF 01-MAR-2002; 2002WO-US006518.  
 PR 01-MAR-2001; 2001US-0272751P.  
 PA (CONF-) CONFORMA THERAPEUTICS CORP.  
 PI Filtz LC, Burrows FU;  
 DR WPI: 2002-698710/75.  
 DR P-PSDB; ABG95083.  
 XX  
 PT Treating genetically-defined disease associated with chromosomal  
 PT aberrations yielding oncogenic fusion proteins, e.g. cell proliferative  
 diseases, involves administering an inhibitor of heat shock protein 90.  
 XX  
 PS Disclosure; Page 246-247; 389PP; English.  
 CC The invention describes a method of treating genetically-defined disease  
 CC associated with chromosomal aberrations yielding oncogenic fusion  
 CC proteins (I), treating cancerous cells containing (I) in a heterogeneous  
 CC cell population, treating proliferative diseases associated with mutant

CC protein or cellular protein isoforms (II) dependent on heat shock protein  
 CC (HSP)-90, or selectively treating cells expressing (II) involving  
 CC administering HSP90-inhibitor. The method is useful for treating  
 CC genetically-defined disease with chromosomal aberration yielding  
 CC oncogenic fusion protein, treating cancerous cells containing fusion  
 CC protein in heterogeneous cell population, treating proliferative disease  
 CC (e.g. rheumatoid arthritis or cancer) associated with mutant protein or  
 CC cellular protein isoform dependent on heat shock protein (HSP)-90 (e.g.  
 CC p53), or selectively treating cells expressing mutant protein or cellular  
 CC protein isoform in a patient heterozygous for (II). The method is useful  
 CC for treating a disease e.g. haematopoietic disorder such as T or B cell  
 CC lymphoma, chronic myeloid leukaemia (CML), APL, ALL, NHL and CMML,  
 CC or a disease characterised by a solid tumour such as papillary thyroid  
 CC carcinoma, Ewing's sarcoma, melanoma, liposarcoma, rhabdomyosarcoma and  
 CC synovial sarcoma. The method is also useful for treating viral  
 CC infections. This represents the DNA sequence of a chromosome aberration  
 XX  
 SQ Sequence 2791 BP; 842 A; 548 C; 625 G; 776 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 2172; DB 6; Length 2791;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 CCCCTGTATGAGTGTGATGATCTTACGAGACGATTTGCTACTTTGGACCTTGATAGCA 60  
 DB 620 CCCCTGTATGAGTGTGATGATCTTACGAGACGATTTGCTACTTTGGACCTTGATAGCA 679  
 QY 61 CTTCCAGAAATGTCACACACGATGAAATATCTCTGCTGAAGACACTGATATAAAGCT 120  
 DB 680 CTTCCAGAAATGTCACACACGATGAAATATCTCTGCTGAAGACACTGATATAAAGCT 739  
 QY 121 CCTCAAGCTCTCTGCTTTTATCTTCAACTCTCACTTTCTTGAAGTTTACAGAAAAA 180  
 DB 740 CCTCAAGCTCTCTGCTTTTATCTTCAACTCTCACTTTCTTGAAGTTTACAGAAAAA 799  
 QY 181 AATTTATATACGACTCTTTTAAAAAGATCTATGCTTGAATAAGAGAAACACAGT 240  
 DB 800 AATTTATATACGACTCTTTTAAAAAGATCTATGCTTGAATAAGAGAAACACAGT 859  
 QY 241 CTGGCCAGGAGAGTGTGCTGCAATTTGATGATTTTAAAGCAATTTGCCCTACTGGA 300  
 DB 860 CTGGCCAGGAGAGTGTGCTGCAATTTGATGATTTTAAAGCAATTTGCCCTACTGGA 919  
 QY 301 ATTAAGAGATGACAGACCTGGAGACCTTAAGATGCAAGCTTTCTTGAAGCTTTA 360  
 DB 920 ATTAAGAGATGACAGACCTGGAGACCTTAAGATGCAAGCTTTCTTGAAGCTTTA 979  
 QY 361 GGTAGATGAGACAGAAAGTATGATCTTAAAAAGCATGATGAGATCAAAATGTTTTA 420  
 DB 980 GGTAGATGAGACAGAAAGTATGATCTTAAAAAGCATGATGAGATCAAAATGTTTTA 1039  
 QY 421 TATCAACATCTTTATATTTGATTCATTTGATTAACAGTGTGTTAGTATGATTTT 480  
 DB 1040 TATCAACATCTTTATATTTGATTCATTTGATTAACAGTGTGTTAGTATGATTTT 1099  
 QY 481 TCTATTTCTTTCCCTTGAAGTATTTCAAGTATCAAAACTCTTCCATGAGGCGATGA 540  
 DB 1100 TCTATTTCTTTCCCTTGAAGTATTTCAAGTATCAAAACTCTTCCATGAGGCGATGA 1159  
 QY 541 TCTATGAGACCTCTTAATGAGATGATCTGGGATGTTGACCCCAACATCTCCCAA 600  
 DB 1160 TCTATGAGACCTCTTAATGAGATGATCTGGGATGTTGACCCCAACATCTCCCAA 1219  
 QY 601 GCATTAATATCCATATGCGCTGTATGTTTAAATCAGACAGAGCATGTTTATGTTG 660  
 DB 1220 GCATTAATATCCATATGCGCTGTATGTTTAAATCAGACAGAGCATGTTTATGTTG 1279  
 QY 661 TACAAAAGAAATGTTATGCTGGGATGAGAGTATGACCATGATGATGATCTTCAA 720  
 DB 1280 TACAAAAGAAATGTTATGCTGGGATGAGAGTATGACCATGATGATGATGATCTTCAA 1339  
 QY 721 GCTACTTATATAGATCTTAAAAATGGGACAGAGCATGTAACAAACACCCATATA 780



Dp	1340	GCTACTTAAATAAGGACTTTAAATATGGGACAGAGAGACTGTGAACAAGACACCTAATAA	1399
Qy	781	TGGGTGATGTCCTGAAGTACGAATCTTCTGGAAACGCAACTCTTTTAAGAGTCCCT	840
Dp	1400	TGGGTGATGTCCTGAAGTACGAATCTTCTGGAAACGCAACTCTTTTAAGAGTCCCT	1459
Qy	841	AATTAGAAACCCACCAACTTCACATATCATTAATATGCAAAACATATGGAAGAGTGT	900
Dp	1460	AATTTAGAACCCACCAACTTCACATATCATTAATATGCAAAACATATGGAAGAGTGT	1519
Qy	901	CTTGAAATGTTGGGAGAGAAAATCTATTTGCTCTGTGGTCTCTTCATCTCAGAAATG	960
Dp	1520	CTTGAAATGTTGGGAGAGAAAATCTATTTGCTCTGTGGTCTCTTCATCTCAGAAATG	1579
Qy	961	CCAAATCAGGTCAGAGTTTGTCTACATTTTGTATGTGTGAAGTCTTCTCCAAAGTATAT	1639
Dp	1580	CCAAATCAGGTCAGAGTTTGTCTACATTTTGTATGTGTGAAGTCTTCTCCAAAGTATAT	
Qy	1021	TAACTATTAAGAGGCTGTGAACAAAACAGATGATTAAGCTGGGAAACCGAGCACAGC	1080
Dp	1640	TAACTATTAAGAGGCTGTGAACAAAACAGATGATTAAGCTGGGAAACCGAGCACAGC	1699
Qy	1081	TCAATGTTCTAGTCTTGAGAGGTTGAGAGGAGATGAGCTTGAAACACAGTGTTCAA	1140
Dp	1700	TCAATGTTCTAGTCTTGAGAGGTTGAGAGGAGATGAGCTTGAAACACAGTGTTCAA	1759
Qy	1141	GGCCAGCTGGGCAACATACAGATCTGTCTCTCAAAAAAAAAAAAAAAAAAGAA	1200
Dp	1760	GGCCAGCTGGGCAACATACAGATCTGTCTCTCTCAAAAAAAAAAAAAAAAAAGAA	1819
Qy	1201	GAGAGAGGGCCGGGCGTGTGGCTTCAAGCCCTGTATCCCAACACTTTGGAGGCCAGAC	1260
Dp	1820	GAGAGAGGGCCGGGCGTGTGGCTTCAAGCCCTGTATCCCAACACTTTGGAGGCCAGAC	1879
Qy	1261	GAGCGGATTCACCTGTGTGTCAGAGTTTGAACCAAGCTGGCCAAACATGGCAAAACCCGT	1320
Dp	1880	GAGCGGATTCACCTGTGTGTCAGAGTTTGAACCAAGCTGGCCAAACATGGCAAAACCCGT	1939
Qy	1321	CTGTATCTCAAAATGCAAAATTTAGCCAGGCGGTGTAGGAGCACTGTATCCAGCTAC	1380
Dp	1940	CTGTATCTCAAAATGCAAAATTTAGCCAGGCGGTGTAGGAGCACTGTATCCAGCTAC	1999
Qy	1381	TTGGAGGCTGAGGAGAGAAATGCTTGAACCCAGAGGTGTGAGAGTTGCACTAAGCTGA	1440
Dp	2000	TTGGAGGCTGAGGAGAGAAATGCTTGAACCCAGAGGTGTGAGAGTTGCACTAAGCTGA	2059
Qy	1441	GATGCTGCGTTGCACTCCAGCTGGGCGACAAAGCAAGAATCTGTCTCAGAAAAAAA	1500
Dp	2060	GATGCTGCGTTGCACTCCAGCTGGGCGACAAAGCAAGAATCTGTCTCAGAAAAAAA	2119
Qy	1501	AAAAAAAAAGAGAGAGAGAAAGAGAAACATATTTGGAGAGAGAGATGGGAGAT	1560
Dp	2120	AAAAAAAAAGAGAGAGAGAAAGAGAAACATATTTGGAGAGAGAGATGGGAGAT	2179
Qy	1561	TGCAAGGAAATTTGTCTTTATTCACAAATGTAAAGAGCCATTAAGGATCCCATTTTG	1620
Dp	2180	TGCAAGGAAATTTGTCTTTATTCACAAATGTAAAGAGCCATTAAGGATCCCATTTTG	2239
Qy	1621	TCTCTTTGGTGTCTAATTTGTCCCTAACAACTGTGTTGACAGTGAAGAAAAATATTGGA	1680
Dp	2240	TCTCTTTGGTGTCTAATTTGTCCCTAACAACTGTGTTGACAGTGAAGAAAAATATTGGA	2299
Qy	1681	ATAACCATATCCCTGTGCGTTATTAACCTAGCAACCTTGCAATGAGATGAGAGATCC	1740
Dp	2300	ATAACCATATCCCTGTGCGTTATTAACCTAGCAACCTTGCAATGAGATGAGAGATCC	2359
Qy	1741	ACAGAGAAACCTTGAATGACAACTGTCTTAATTTATCTTAATGTAATAGTTGTAAA	1800
Dp	2360	ACAGAGAAACCTTGAATGACAACTGTCTTAATTTATCTTAATGTAATAGTTGTAAA	2419
Qy	1801	AGAGTTAAATATGTTACTTCACTGAATCATTAATTTAATATTTGGGTATAG	1860
Dp	2420	AGAGTTAAATATGTTACTTCACTGAATCATTAATTTAATATTTGGGTATAG	2479

QY	1861	ATTTTATTATTAACATGATTTCTCTTTCGTGAATATTTGAAATGAGCTCTAAAGCTTCATA	1982
Db	2480	ATTTTTTATTAACAATGATTTCTCTTTCGTGAATATTTGAAATGAGCTCTAAAGCTTCATA	2533
QY	1921	AATTTATTAACCTTAGAATAATGATTCCTAATTAACAACGATGTGTAATGTGTAACATGCAATAT	1983
Db	2540	AATTTTAACTTAGAATAATGATTCCTAATTAACAACGATGTGTAATGTGTAACATGCAATAT	2593
QY	1981	GGGCTACGACAGCCATTTCTCTTGATTTTATGTAACCTTTATGACGCAATTTGGCTTC	2041
Db	2600	GGGCTACGACAGCCATTTCTCTTGATTTTATGTAACCTTTATGACGCAATTTGGCTTC	2653
QY	2041	TGGCTCACTTTCAATCAGTTAATTAATGATTAATATTTTGGAGCTGTGAAGATTAATA	2103
Db	2660	TGGCTCACTTTCAATCAGTTAATTAATGATTAATATTTTGGAGCTGTGAAGATTAATA	2711
QY	2101	TACCAATTAATTAATTAATTAATGAAGATTTAATGAAGTTAATGAAGTTAATTAATTAATTAATGAAGTTAAT	2163
Db	2720	TACCAATTAATTAATTAATTAATGAAGATTTAATGAAGTTAATGAAGTTAATTAATTAATTAATGAAGTTAAT	2773
QY	2161	GGAATTAACCTTG 2172	
Db	2780	GGAATTAACCTTG 2791	
RESULT 3			
ABS73288			
ID	ABS73288	standard; DNA; 2791 BP.	
AC	ABS73288;		
DT	04-DEC-2002	(first entry)	
XX			
DE		DNA encoding human translocation del(12p) protein #3.	
XX			
XX		Chromosome aberration; oncogenic fusion protein; cancer;	
KW		proliferative disease; cellular protein isoform; heat shock protein 90;	
KW		HSP-90; rheumatoid arthritis; cancer; haematopoietic disorder;	
KM		T cell lymphoma; B cell lymphoma; chronic myeloid leukaemia; CML;	
KM		acute myeloid leukaemia; AML; chronic myelomonocytic leukaemia; CMML;	
KW		acute lymphoblastic leukaemia; ALL; APL; NHL; solid tumour;	
KW		papillary thyroid carcinoma; Ewing's sarcoma; melanoma; liposarcoma;	
KW		rhabdomyosarcoma; synovial sarcoma; viral infection; gene; ds.	
XX			
OS		Homo sapiens.	
XX			
PN		MO200269900-A2.	
XX			
PD		12-SEP-2002.	
XX			
PF		01-MAR-2002; 2002W0-US006518.	
XX			
PR		01-MAR-2001; 2001US-0272751P.	
XX			
PA		(CONF-) CONFORNA THERAPEUTICS CORP.	
XX			
PI		Fritz LC; Burrows FU;	
XX			
DR		WPI; 2002-698710/75.	
XX		P-PSDB; ABG95084.	
XX			
PT		Treating genetically-defined disease associated with chromosomal	
XX		aberrations yielding oncogenic fusion proteins, e.g. cell proliferative	
CC		diseases, involves administering an inhibitor of heat shock protein 90.	
XX			
PS		Disclosure; Page 248-249; 389pp; English.	
XX			
CC		The invention describes a method of treating genetically-defined disease	
XX		associated with chromosomal aberrations yielding oncogenic fusion	
CC		proteins (I), treating cancerous cells containing (I) in a heterogeneous	
CC		cell population, treating proliferative diseases associated with mutant	
CC		protein or cellular protein isoforms (II) dependent on heat shock protein	



CC (HSP)-90, or selectively treating cells expressing (ii) involving  
CC administering HSP90-inhibitor. The method is useful for treating  
CC genetically-defined disease with chromosomal aberration yielding  
CC oncogenic fusion protein, treating cancerous cells containing fusion  
CC protein in heterogeneous cell population, treating proliferative disease  
CC (e.g. rheumatoid arthritis or cancer) associated with mutant protein or  
CC cellular protein isoform dependent on heat shock protein (HSP)-90 (e.g.  
CC p53), or selectively treating cells expressing mutant protein or cellular  
CC protein isoform in a patient heterozygous for (ii). The method is useful  
CC for treating a disease e.g. haematopoietic disorder such as T or B cell  
CC lymphoma, chronic myeloid leukaemia (CML), APL, ALL, AML, NHL and CML,  
CC or a disease characterised by a solid tumour such as papillary thyroid  
CC carcinoma, Ewing's sarcoma, melanoma, liposarcoma, rhabdomyosarcoma and  
CC synovial sarcoma. The method is also useful for treating viral  
CC infections. This represents the DNA sequence of a chromosome aberration  
XX  
SQ Sequence 2791 BP; 842 A; 548 C; 625 G; 776 T; 0 U; 0 Other;

Query Match 100.0%; Score 2172; DB 6; Length 2791;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCCTGATAGAGTGTAGTACTTACAGACGCAATTCGTAATTTGGAGCTTTGATAGCA 60  
DB CCCCTGATAGAGTGTAGTACTTACAGACGCAATTCGTAATTTGGAGCTTTGATAGCA 679  
QY 61 CTTCAGAGATGTCAACACGATGAATATCTCTGCTGAAGACAGTGGATAAAAACAGT 120  
DB CTTCAGAGATGTCAACACGATGAATATCTCTGCTGAAGACAGTGGATAAAAACAGT 739  
QY 121 CCTTAAAGCTCTCTGTTTTTATCTTCAACTCTCACTTTCTTGAATTTACAGAAAA 180  
DB CCTTAAAGCTCTCTGTTTTTATCTTCAACTCTCACTTTCTTGAATTTACAGAAAA 799  
QY 740 CCTTAAAGCTCTCTGTTTTTATCTTCAACTCTCACTTTCTTGAATTTACAGAAAA 799  
DB CCTTAAAGCTCTCTGTTTTTATCTTCAACTCTCACTTTCTTGAATTTACAGAAAA 799  
QY 181 ATATTTATATACGACTCTTTAAAAAGATCTATGTCTTGAATAAGAGAACACAGGT 240  
DB ATATTTATATACGACTCTTTAAAAAGATCTATGTCTTGAATAAGAGAACACAGGT 859  
QY 800 ATATTTATATACGACTCTTTAAAAAGATCTATGTCTTGAATAAGAGAACACAGGT 859  
DB ATATTTATATACGACTCTTTAAAAAGATCTATGTCTTGAATAAGAGAACACAGGT 859  
QY 241 CTGGCCAGAGGAGCTGCTGCAATTTGGTGCAGTTTGAATGCAACATGTCCCTACTGGA 300  
DB CTGGCCAGAGGAGCTGCTGCAATTTGGTGCAGTTTGAATGCAACATGTCCCTACTGGA 919  
QY 860 CTGGCCAGAGGAGCTGCTGCAATTTGGTGCAGTTTGAATGCAACATGTCCCTACTGGA 919  
DB CTGGCCAGAGGAGCTGCTGCAATTTGGTGCAGTTTGAATGCAACATGTCCCTACTGGA 919  
QY 301 ATAAAGAGACTGCAAGACCTGGAGACATCTTAAAGTGCACAGTTTCTTGAATTTA 360  
DB ATAAAGAGACTGCAAGACCTGGAGACATCTTAAAGTGCACAGTTTCTTGAATTTA 979  
QY 920 ATAAAGAGACTGCAAGACCTGGAGACATCTTAAAGTGCACAGTTTCTTGAATTTA 979  
DB ATAAAGAGACTGCAAGACCTGGAGACATCTTAAAGTGCACAGTTTCTTGAATTTA 979  
QY 361 GGTAGAGTAGAGAGAGAGTATCTTAAAGAGATGGTAGAGAGATCAATGTTTTA 420  
DB GGTAGAGTAGAGAGAGAGTATCTTAAAGAGATGGTAGAGAGATCAATGTTTTA 1039  
QY 980 GGTAGAGTAGAGAGAGAGTATCTTAAAGAGATGGTAGAGAGATCAATGTTTTA 1039  
DB GGTAGAGTAGAGAGAGAGTATCTTAAAGAGATGGTAGAGAGATCAATGTTTTA 1039  
QY 421 TATCAACATCTTTATTTATTTGATTCATTGATTAACAGTGTGTTAGATAGATTTT 480  
DB TATCAACATCTTTATTTATTTGATTCATTGATTAACAGTGTGTTAGATAGATTTT 1099  
QY 1040 TATCAACATCTTTATTTATTTGATTCATTGATTAACAGTGTGTTAGATAGATTTT 1099  
DB TATCAACATCTTTATTTATTTGATTCATTGATTAACAGTGTGTTAGATAGATTTT 1099  
QY 481 TCTATTTCTTTTCCCTTGAAGCTTTCTTCAAGTAAACAACACTCTTCATGAGGCCATGA 540  
DB TCTATTTCTTTTCCCTTGAAGCTTTCTTCAAGTAAACAACACTCTTCATGAGGCCATGA 1159  
QY 1100 TCTATTTCTTTTCCCTTGAAGCTTTCTTCAAGTAAACAACACTCTTCATGAGGCCATGA 1159  
DB TCTATTTCTTTTCCCTTGAAGCTTTCTTCAAGTAAACAACACTCTTCATGAGGCCATGA 1159  
QY 541 TCTATTTCTTTTCCCTTGAAGCTTTCTTCAAGTAAACAACACTCTTCATGAGGCCATGA 600  
DB TCTATTTCTTTTCCCTTGAAGCTTTCTTCAAGTAAACAACACTCTTCATGAGGCCATGA 600  
QY 1160 TCTATTTCTTTTCCCTTGAAGCTTTCTTCAAGTAAACAACACTCTTCATGAGGCCATGA 1219  
DB TCTATTTCTTTTCCCTTGAAGCTTTCTTCAAGTAAACAACACTCTTCATGAGGCCATGA 1219  
QY 601 GCATTAATATCCATCATGCGCTGATGTTTAAATCAGAGAGAGATGTTTAAATGTTTG 660  
DB GCATTAATATCCATCATGCGCTGATGTTTAAATCAGAGAGAGATGTTTAAATGTTTG 1279  
QY 1220 GCATTAATATCCATCATGCGCTGATGTTTAAATCAGAGAGAGATGTTTAAATGTTTG 1279  
DB GCATTAATATCCATCATGCGCTGATGTTTAAATCAGAGAGAGATGTTTAAATGTTTG 1279  
QY 661 TACAAAAGAGAGATGTTTAAATGAGAGAGAGATGAGATAGACCATGTCACATTTCA 720  
DB TACAAAAGAGAGATGTTTAAATGAGAGAGAGATGAGATAGACCATGTCACATTTCA 1339  
QY 1280 TACAAAAGAGAGATGTTTAAATGAGAGAGAGATGAGATAGACCATGTCACATTTCA 1339  
DB TACAAAAGAGAGATGTTTAAATGAGAGAGAGATGAGATAGACCATGTCACATTTCA 1339  
QY 721 GCTACTTTAATAAGAGATCTTAAATGGGAGAGAGACTGTGACAAAGACCCCTAATA 780  
DB GCTACTTTAATAAGAGATCTTAAATGGGAGAGAGACTGTGACAAAGACCCCTAATA 780  
QY 1340 GCTACTTTAATAAGAGATCTTAAATGGGAGAGAGACTGTGACAAAGACCCCTAATA 1399  
DB GCTACTTTAATAAGAGATCTTAAATGGGAGAGAGACTGTGACAAAGACCCCTAATA 1399

QY 781 TGGGTGATGTCTGAGTAGCAAAATCTTCTGGAAAAGCAAACTCTTTAAGAGAGTCCCT 840  
DB TGGGTGATGTCTGAGTAGCAAAATCTTCTGGAAAAGCAAACTCTTTAAGAGAGTCCCT 1459  
QY 841 AATTGAAAACACCCCAAACTTACATATATTTAGCAAAACAATTGAAAGAAAGTTG 900  
DB AATTGAAAACACCCCAAACTTACATATATTTAGCAAAACAATTGAAAGAAAGTTG 1519  
QY 1460 AATTGAAAACACCCCAAACTTACATATATTTAGCAAAACAATTGAAAGAAAGTTG 1519  
DB AATTGAAAACACCCCAAACTTACATATATTTAGCAAAACAATTGAAAGAAAGTTG 1519  
QY 901 CTGGAATGTTGGGAGAGAGAAAATCTATTTGCTCTGTGGTCTCTTCACTTCAAGAAAG 960  
DB CTGGAATGTTGGGAGAGAGAAAATCTATTTGCTCTGTGGTCTCTTCACTTCAAGAAAG 1579  
QY 1520 CTGGAATGTTGGGAGAGAGAAAATCTATTTGCTCTGTGGTCTCTTCACTTCAAGAAAG 1579  
DB CTGGAATGTTGGGAGAGAGAAAATCTATTTGCTCTGTGGTCTCTTCACTTCAAGAAAG 1579  
QY 961 CCAATCAGGTCAGAGTTTGCTACATTTTATGTTGTTGATGATGCTTCCCAAGGATAT 1020  
DB CCAATCAGGTCAGAGTTTGCTACATTTTATGTTGTTGATGATGCTTCCCAAGGATAT 1639  
QY 1580 CCAATCAGGTCAGAGTTTGCTACATTTTATGTTGTTGATGATGCTTCCCAAGGATAT 1639  
DB CCAATCAGGTCAGAGTTTGCTACATTTTATGTTGTTGATGATGCTTCCCAAGGATAT 1639  
QY 1021 TAACTATATAGAGAGTGTGACAAAACAGAAATGATTAAGTGCAGAACCGTGGACACGC 1080  
DB TAACTATATAGAGAGTGTGACAAAACAGAAATGATTAAGTGCAGAACCGTGGACACGC 1699  
QY 1640 TAACTATATAGAGAGTGTGACAAAACAGAAATGATTAAGTGCAGAACCGTGGACACGC 1699  
DB TAACTATATAGAGAGTGTGACAAAACAGAAATGATTAAGTGCAGAACCGTGGACACGC 1699  
QY 1081 TCATAGTTCTAGCTGCTTGGAGGTTGAGAGAGGAGATGCTTGAACACAGGTGTTCA 1140  
DB TCATAGTTCTAGCTGCTTGGAGGTTGAGAGAGGAGATGCTTGAACACAGGTGTTCA 1759  
QY 1700 TCATAGTTCTAGCTGCTTGGAGGTTGAGAGAGGAGATGCTTGAACACAGGTGTTCA 1759  
DB TCATAGTTCTAGCTGCTTGGAGGTTGAGAGAGGAGATGCTTGAACACAGGTGTTCA 1759  
QY 1141 GGCACGCTGGGACATATACAAAGATCTCTCTCAAAAAAAAAAAAAAAAAAGAA 1200  
DB GGCACGCTGGGACATATACAAAGATCTCTCTCAAAAAAAAAAAAAAAAAAGAA 1819  
QY 1760 GGCACGCTGGGACATATACAAAGATCTCTCTCAAAAAAAAAAAAAAAAAAGAA 1819  
DB GGCACGCTGGGACATATACAAAGATCTCTCTCAAAAAAAAAAAAAAAAAAGAA 1819  
QY 1201 GAGAGAGGGCCGGGCGTGGTGGCTCAGCGCTGTAATCCAGAGACTTTGGAGAGCGAGCC 1260  
DB GAGAGAGGGCCGGGCGTGGTGGCTCAGCGCTGTAATCCAGAGACTTTGGAGAGCGAGCC 1879  
QY 1820 GAGAGAGGGCCGGGCGTGGTGGCTCAGCGCTGTAATCCAGAGACTTTGGAGAGCGAGCC 1879  
DB GAGAGAGGGCCGGGCGTGGTGGCTCAGCGCTGTAATCCAGAGACTTTGGAGAGCGAGCC 1879  
QY 1261 GGGCGGATCACTGTGTGAGAGAGTTGAGACCAAGCTTGGCAACATGGCAAAACCCCGT 1320  
DB GGGCGGATCACTGTGTGAGAGAGTTGAGACCAAGCTTGGCAACATGGCAAAACCCCGT 1939  
QY 1880 GGGCGGATCACTGTGTGAGAGAGTTGAGACCAAGCTTGGCAACATGGCAAAACCCCGT 1939  
DB GGGCGGATCACTGTGTGAGAGAGTTGAGACCAAGCTTGGCAACATGGCAAAACCCCGT 1939  
QY 1321 CTGTACTCAAAATGCAAAAATTTAGCGAGCGTGTGTAGCGACGCTGTATCTCCAGCTAC 1380  
DB CTGTACTCAAAATGCAAAAATTTAGCGAGCGTGTGTAGCGACGCTGTATCTCCAGCTAC 1999  
QY 1940 CTGTACTCAAAATGCAAAAATTTAGCGAGCGTGTGTAGCGACGCTGTATCTCCAGCTAC 1999  
DB CTGTACTCAAAATGCAAAAATTTAGCGAGCGTGTGTAGCGACGCTGTATCTCCAGCTAC 1999  
QY 1381 TTGGAGGCTGAGGCGAGAGAAATCGCTGAACCCAGAGAGTGGAGGTTGACATTAAGCTGA 1440  
DB TTGGAGGCTGAGGCGAGAGAAATCGCTGAACCCAGAGAGTGGAGGTTGACATTAAGCTGA 1440  
QY 1441 GATCGTGCCTTGCATCCAGCCTGGGCGCAAGAGCAAGACTGTCTCAGAAAAAATA 1500  
DB GATCGTGCCTTGCATCCAGCCTGGGCGCAAGAGCAAGACTGTCTCAGAAAAAATA 2059  
QY 2060 GATCGTGCCTTGCATCCAGCCTGGGCGCAAGAGCAAGACTGTCTCAGAAAAAATA 2059  
DB GATCGTGCCTTGCATCCAGCCTGGGCGCAAGAGCAAGACTGTCTCAGAAAAAATA 2059  
QY 1501 AAAAAAAGAGAGAGAGAGAAAGAAACAATTTTGGAGAGAAAGATGGGAGAAAT 1560  
DB AAAAAAAGAGAGAGAGAGAAAGAAACAATTTTGGAGAGAAAGATGGGAGAAAT 2179  
QY 2120 AAAAAAAGAGAGAGAGAGAAAGAAACAATTTTGGAGAGAAAGATGGGAGAAAT 2179  
DB AAAAAAAGAGAGAGAGAGAAAGAAACAATTTTGGAGAGAAAGATGGGAGAAAT 2179  
QY 1561 TGCAGAGAAATGTGCTTTATGCAACAAATGTGAAGACCAATAAGGATCCCTATTG 1620  
DB TGCAGAGAAATGTGCTTTATGCAACAAATGTGAAGACCAATAAGGATCCCTATTG 2239  
QY 2180 TGCAGAGAAATGTGCTTTATGCAACAAATGTGAAGACCAATAAGGATCCCTATTG 2239  
DB TGCAGAGAAATGTGCTTTATGCAACAAATGTGAAGACCAATAAGGATCCCTATTG 2239  
QY 1621 TCTCTTTTGGTGTCTATTTGTCCCTTAAACAAGTCTTTGACAGTGAAGAAAAATATTCAGA 1680  
DB TCTCTTTTGGTGTCTATTTGTCCCTTAAACAAGTCTTTGACAGTGAAGAAAAATATTCAGA 2299  
QY 2240 TCTCTTTTGGTGTCTATTTGTCCCTTAAACAAGTCTTTGACAGTGAAGAAAAATATTCAGA 2299  
DB TCTCTTTTGGTGTCTATTTGTCCCTTAAACAAGTCTTTGACAGTGAAGAAAAATATTCAGA 2299  
QY 1681 ATAAACATATCCCTGTGCGCTTATTAACCTAGCAACCTTGAATGAATGAGAGATCC 1740  
DB ATAAACATATCCCTGTGCGCTTATTAACCTAGCAACCTTGAATGAATGAGAGATCC 2359  
QY 2300 ATAAACATATCCCTGTGCGCTTATTAACCTAGCAACCTTGAATGAATGAGAGATCC 2359  
DB ATAAACATATCCCTGTGCGCTTATTAACCTAGCAACCTTGAATGAATGAGAGATCC 2359  
QY 1741 ACAGAGAAACTGATGACACAACTGCTTATTTAATCTTATTTGACATTAAGTTGTAAA 1800  
DB ACAGAGAAACTGATGACACAACTGCTTATTTAATCTTATTTGACATTAAGTTGTAAA 2419  
QY 2360 ACAGAGAAACTGATGACACAACTGCTTATTTAATCTTATTTGACATTAAGTTGTAAA 2419  
DB ACAGAGAAACTGATGACACAACTGCTTATTTAATCTTATTTGACATTAAGTTGTAAA 2419  
QY 1801 AGAGTTAAATAATGTACTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1860  
DB AGAGTTAAATAATGTACTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2479  
QY 2420 AGAGTTAAATAATGTACTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2479  
DB AGAGTTAAATAATGTACTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2479







Dp	1283	TACAAAAGAAAGATTGTATGGGTGGGGATGGAGGTATAGACATGCAATGGTCACTTTCAA	1342
Qy	721	GCTACTTTAATAAAGATCTTTAAATGGGACAGGAGACTGTGGAAACAAGACACCTTATAA	780
Dp	1343	GCATCTTTAATAAAGATCTTTAAATGGGACAGGAGACTGTGGAAACAAGACACCTTATAA	1402
Qy	781	TGGGTGATGTCTGAAGTAGCAAACTCTTCTGGAAGCCCAACTCTTTTAAGGAAGTCCCT	840
Dp	1403	TGGGTGATGTCTGAAGTAGCAAACTCTTCTGGAAGCCCAACTCTTTTAAGGAAGTCCCT	1462
Qy	841	AATTTAGAAAACCCCAAACTTCACATATCATAATTAGCAACAATTTGAAGAGAGTTGG	900
Dp	1463	AATTTAGAAAACCCCAAACTTCACATATCATAATTAGCAACAATTTGAAGAGAGTTGG	1522
Qy	901	CTTGAAGTTGGGAGAGGAAAAACTTTTGGCTTCGTGGGTCTCTCACTCAGAAATG	960
Dp	1523	CTTGAAGTTGGGAGAGGAAAAACTTTTGGCTTCGTGGGTCTCTCACTCAGAAATG	1582
Qy	961	CCATCAGGCTCAAGGTTTGCTACATTTTGTATGTGTGAATGCTTCTCCCAAGGTTAT	1020
Dp	1583	CCATCAGGCTCAAGGTTTGCTACATTTTGTATGTGTGAATGCTTCTCCCAAGGTTAT	1642
Qy	1021	TAACTATATAGAGAGTTGTGACAAAAACAAGATGATAAAGCTGGAAACGTGGACACGC	1080
Dp	1643	TAACTATATAGAGAGTTGTGACAAAAACAAGATGATAAAGCTGGAAACGTGGACACGC	1702
Qy	1081	TCATATGTTCTTAGCTGCTTTGGAGGTTGAGAGGAGAGATGGCTTGAACAACAGTGTTCAA	1140
Dp	1703	TCATATGTTCTTAGCTGCTTTGGAGGTTGAGAGGAGAGATGGCTTGAACAACAGTGTTCAA	1762
Qy	1141	GGCCAGCTGGGGCAACATPACAAATCTGTCTCAAAAAAATAAAAAAATAAAAAAATA	1200
Dp	1763	GGCCAGCTGGGGCAACATPACAAATCTGTCTCAAAAAAATAAAAAAATAAAAAAATA	1822
Qy	1201	GAGAGAGGCGCGGCGTGTGGCTCACGCGCTGTATCCACACTTTTGGAGGCGGAGCC	1260
Dp	1823	GAGAGAGGCGCGGCGTGTGGCTCACGCGCTGTATCCACACTTTTGGAGGCGGAGCC	1882
Qy	1261	GGGCGGATCACCTGTGTGACAGAGTTTGAGACCAAGCTTGGCAATGGCAAAACCCCGT	1320
Dp	1883	GGGCGGATCACCTGTGTGACAGAGTTTGAGACCAAGCTTGGCAATGGCAAAACCCCGT	1942
Qy	1321	CTGTACTCAAAATTCAAAAATTTACCCAGCGCTGGTACAGGCACTGTAAATCCAGGTAC	1380
Dp	1943	CTGTACTCAAAATTCAAAAATTTACCCAGCGCTGGTACAGGCACTGTAAATCCAGGTAC	2002
Qy	1381	TTGGGAGGCTGAGGACAGAGAAATGCTTGAACCCAGAGGTTGAGGTTGCACTAAGCTGA	1440
Dp	2003	TTGGGAGGCTGAGGACAGAGAAATGCTTGAACCCAGAGGTTGAGGTTGCACTAAGCTGA	2062
Qy	1441	GATGTCGCCGTGGCACTCCAGCTGGGCGACAAAGGCAAGACTGTGTCTCGAATAAAAAA	1500
Dp	2063	GATGTCGCCGTGGCACTCCAGCTGGGCGACAAAGGCAAGACTGTGTCTCGAATAAAAAA	2122
Qy	1501	AAAAAAGAGAGAGAGAGAAAGAAACAATTTTGGGAGAGAAAGATGGGGAAGCAT	1560
Dp	2123	AAAAAAGAGAGAGAGAGAAAGAAACAATTTTGGGAGAGAAAGATGGGGAAGCAT	2182
Qy	1561	TGCAAGGAAATTTGCTTTATTCACAAATAATGTAAAGGACCAATTAAGGATCCCTATTGG	1620
Dp	2183	TGCAAGGAAATTTGCTTTATTCACAAATAATGTAAAGGACCAATTAAGGATCCCTATTGG	2242
Qy	1621	TCTCTTTTGTGTCTATTTGTCCCTAAACACTGCTTTGACAGTGAAGAAAAATTCAGA	1680
Dp	2243	TCTCTTTTGTGTCTATTTGTCCCTAAACACTGCTTTGACAGTGAAGAAAAATTCAGA	2302
Qy	1681	ATTAACCATATCCCTGTGCGCTTATTTACTTAGCAACCTTGGCAATGAAGATGACAGATCC	1740
Dp	2303	ATTAACCATATCCCTGTGCGCTTATTTACTTAGCAACCTTGGCAATGAAGATGACAGATCC	2362
Qy	1741	ACAGAGAAACCTGAATGCAACAAGTCTATTTTAAATCTTATTTGACATAAAGTTGTAAA	1800
Dp	2363	ACAGAGAAACCTGAATGCAACAAGTCTATTTTAAATCTTATTTGACATAAAGTTGTAAA	2422

QY	1801	AGAGTAAATATGTCTTCATGATTCATTTATATTTATATTTATTTGCGCTAATG	1860
Db	2423	AGAGTTAAATATGTCTTCATGATTCATTTATATTTATATTTTGGCGCTAATG	2482
QY	1861	ATTTTATTTAACAATGATTTCCTTTCTGATATATTTGAATTCGAAGTCTCAAGCTTATA	1920
Db	2483	ATTTTATTTAACAATGATTTCCTTTCTGATATTTGAATTCGAAGTCTCAAGCTTATA	2542
QY	1921	AATTTAATCTTAAAGAAATGATTCATTAACAACGATGTAATTTGTAAACATTCGAGTAAT	1980
Db	2543	AATTTAATCTTAAAGAAATGATTCATTAACAACGATGTAATTTGTAAACATTCGAGTAAT	2602
QY	1981	GGTGTCAAGAACCATTTCTCTTGATTTTAAAGTAAACCTTTATGACAGCAAAATTTGCTTC	2040
Db	2603	GGTGTCTAAGAACCATTTCTCTTGATTTTAAAGTAAACCTTTATGACAGCAAAATTTGCTTC	2662
QY	2041	TGCGTCACTTTCAATCAGTTAAATTAATGATTAATAATTTTGGAACTGTGAGATTAATA	2100
Db	2663	TGCGTCACTTTCAATCAGTTAAATTAATGATTAATAATTTTGGAACTGTGAGATTAATA	2722
QY	2101	TACCAATTAATAATATTAATAAGATTTATATGAAGTTAAATTAATAATTCAGTATGAT	2160
Db	2723	TACCAATTAATAATATTAATAAGATTTATATGAAGTTAAATTAATAATTCAGTATGAT	2782
QY	2161	GGAATTAACCTTG	2172
Db	2783	GGAATTAACCTTG	2794

[illegible]















CC associated with chromosomal aberrations yielding oncogenic fusion  
CC proteins (I), treating cancerous cells containing (II) in a heterogenous  
CC cell population, treating proliferative diseases associated with mutant  
CC protein or cellular protein isoforms (II) dependent on heat shock protein  
CC (HSP)-90, or selectively treating cells expressing (II) involving  
CC administering HSP90-inhibitor. The method is useful for treating  
CC genetically-defined disease with chromosomal aberration yielding  
CC oncogenic fusion protein, treating cancerous cells containing fusion  
CC protein in heterogenous cell population, treating proliferative disease  
CC (e.g. rheumatoid arthritis or cancer) associated with mutant protein or  
CC cellular protein isoform dependent on heat shock protein (HSP)-90 (e.g.  
CC p53), or selectively treating cells expressing mutant protein or cellular  
CC protein isoform in a patient heterozygous for (II). The method is useful  
CC for treating a disease e.g. haematopoietic disorder such as T or B cell  
CC lymphoma, chronic myeloid leukaemia (CML), APL, ALL, NHL and CML,  
CC or a disease characterized by a solid tumour such as papillary thyroid  
CC carcinoma, Ewing's sarcoma, melanoma, liposarcoma, rhabdomyosarcoma and  
CC synovial sarcoma. The method is also useful for treating viral  
CC infections. This represents the DNA sequence of a chromosome aberration  
XX

Sequence 11204 BP; 3305 A; 2273 C; 2373 G; 3253 T; 0 U; 0 Other;

Query Match 100.0%; Score 2172; DB 6; Length 11204;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCCTGATAGAGTTGATGACTTACGAGACGCAATTTGCTACTTTGGGACCTTTGATACGA 60  
DB CCCCTGATAGAGTTGATGACTTACGAGACGCAATTTGCTACTTTGGGACCTTTGATACGA 9015  
QY 61 CTTCCAGGAATGTCAACACAGATGAATATCTGCTGAGACAGTGGATATAAAAGT 120  
DB CTTCCAGGAATGTCAACACAGATGAATATCTGCTGAGACAGTGGATATAAAAGT 9075  
QY 121 CCTCAAGTCTCTCTGTTTATTTCTTCAACTCTCACTTCTTAGAGTTTACAGAAAA 180  
DB CCTCAAGTCTCTCTGTTTATTTCTTCAACTCTCACTTCTTAGAGTTTACAGAAAA 9135  
QY 9076 CTTTCAAGTCTCTCTGTTTATTTCTTCAACTCTCACTTCTTAGAGTTTACAGAAAA 9135  
DB CTTTCAAGTCTCTCTGTTTATTTCTTCAACTCTCACTTCTTAGAGTTTACAGAAAA 9135  
QY 181 ATATTATATACGACTCTTTTAAAAAGATCTATGCTTGAATAAGAGAGACACAGT 240  
DB ATATTATATACGACTCTTTTAAAAAGATCTATGCTTGAATAAGAGAGACACAGT 9195  
QY 9136 ATATTATATACGACTCTTTTAAAAAGATCTATGCTTGAATAAGAGAGACACAGT 9195  
DB ATATTATATACGACTCTTTTAAAAAGATCTATGCTTGAATAAGAGAGACACAGT 9195  
QY 241 CTGGCAGGAGACGCTCTGCAATGTGTGAGATTGGAATGCAACTGTCCCTTACTGGGA 300  
DB CTGGCAGGAGACGCTCTGCAATGTGTGAGATTGGAATGCAACTGTCCCTTACTGGGA 9255  
QY 9196 CTGGCAGGAGACGCTCTGCAATGTGTGAGATTGGAATGCAACTGTCCCTTACTGGGA 9255  
DB CTGGCAGGAGACGCTCTGCAATGTGTGAGATTGGAATGCAACTGTCCCTTACTGGGA 9255  
QY 301 ATTAAGAGACGCTGAGGACCTGGAGACCTTAAGTGCACAGTTTCTATGACTTTA 360  
DB ATTAAGAGACGCTGAGGACCTGGAGACCTTAAGTGCACAGTTTCTATGACTTTA 9256  
QY 9256 ATTAAGAGACGCTGAGGACCTGGAGACCTTAAGTGCACAGTTTCTATGACTTTA 9315  
DB ATTAAGAGACGCTGAGGACCTGGAGACCTTAAGTGCACAGTTTCTATGACTTTA 9315  
QY 361 GGTAGATGAGAGCAGAGATGATCCATAAAGCATGTGAGAGATCAAAATGTTTTA 420  
DB GGTAGATGAGAGCAGAGATGATCCATAAAGCATGTGAGAGATCAAAATGTTTTA 9316  
QY 9316 GGTAGATGAGAGCAGAGATGATCCATAAAGCATGTGAGAGATCAAAATGTTTTA 9375  
DB GGTAGATGAGAGCAGAGATGATCCATAAAGCATGTGAGAGATCAAAATGTTTTA 9375  
QY 421 TATCAACATCCTTTATTTATTTGATTCATTTGATTAACAAGTGTGTTAGTAGATTTT 480  
DB TATCAACATCCTTTATTTATTTGATTCATTTGATTAACAAGTGTGTTAGTAGATTTT 9376  
QY 9376 TATCAACATCCTTTATTTATTTGATTCATTTGATTAACAAGTGTGTTAGTAGATTTT 9435  
DB TATCAACATCCTTTATTTATTTGATTCATTTGATTAACAAGTGTGTTAGTAGATTTT 9435  
QY 481 TCTATCTTTTCCCTTGACGTTTACTTTCAGATTAACAAACTCTTCCATCAGGCCATGA 540  
DB TCTATCTTTTCCCTTGACGTTTACTTTCAGATTAACAAACTCTTCCATCAGGCCATGA 9436  
QY 9436 TCTATCTTTTCCCTTGACGTTTACTTTCAGATTAACAAACTCTTCCATCAGGCCATGA 9495  
DB TCTATCTTTTCCCTTGACGTTTACTTTCAGATTAACAAACTCTTCCATCAGGCCATGA 9495  
QY 541 TCTATAGACCTCCTTAATGAGAGATCTGGGTGATGTGACCCCAAAACATCTCCAAA 600  
DB TCTATAGACCTCCTTAATGAGAGATCTGGGTGATGTGACCCCAAAACATCTCCAAA 9496  
QY 9496 TCTATAGACCTCCTTAATGAGAGATCTGGGTGATGTGACCCCAAAACATCTCCAAA 9555  
DB TCTATAGACCTCCTTAATGAGAGATCTGGGTGATGTGACCCCAAAACATCTCCAAA 9555  
QY 601 GCATTATATCCAAATCATGCGCTGATGTTTAAATCAGCAGAGACATGTTTATGTTTG 660  
DB GCATTATATCCAAATCATGCGCTGATGTTTAAATCAGCAGAGACATGTTTATGTTTG 9556  
QY 9556 GCATTATATCCAAATCATGCGCTGATGTTTAAATCAGCAGAGACATGTTTATGTTTG 9615  
DB GCATTATATCCAAATCATGCGCTGATGTTTAAATCAGCAGAGACATGTTTATGTTTG 9615  
QY 661 TACAAAGAGAGATTGTTATGGGTGGGAGTGAAGTATGACATGATGTCACCTTCAA 720  
DB TACAAAGAGAGATTGTTATGGGTGGGAGTGAAGTATGACATGATGTCACCTTCAA 9616  
QY 9616 TACAAAGAGAGATTGTTATGGGTGGGAGTGAAGTATGACATGATGTCACCTTCAA 9675  
DB TACAAAGAGAGATTGTTATGGGTGGGAGTGAAGTATGACATGATGTCACCTTCAA 9675

QY 721 GCTACTTTATATAAGATCTTAAATGGGACGAGAGCTGTGACAAACACCTTAATTA 780  
DB GCTACTTTATATAAGATCTTAAATGGGACGAGAGAGCTGTGACAAACACCTTAATTA 9735  
QY 781 TGGGTGATGTCTGAATGACAAATCTTCTGGAACGCAACTCTTTAAGAGTCCCT 840  
DB TGGGTGATGTCTGAATGACAAATCTTCTGGAACGCAACTCTTTAAGAGTCCCT 9736  
QY 9736 TGGGTGATGTCTGAATGACAAATCTTCTGGAACGCAACTCTTTAAGAGTCCCT 9795  
DB TGGGTGATGTCTGAATGACAAATCTTCTGGAACGCAACTCTTTAAGAGTCCCT 9795  
QY 841 AATTTAGAACCCCAACCACTTCAATATCAATAATTGCAACAAATGGAAGAGTGG 900  
DB AATTTAGAACCCCAACCACTTCAATATCAATAATTGCAACAAATGGAAGAGTGG 9796  
QY 9796 AATTTAGAACCCCAACCACTTCAATATCAATAATTGCAACAAATGGAAGAGTGG 9855  
DB AATTTAGAACCCCAACCACTTCAATATCAATAATTGCAACAAATGGAAGAGTGG 9855  
QY 901 CTTGAATGTTGGGAGAGAGAAATATATGAGCTCGTGGGTCTCTTCAATCAGAAATG 960  
DB CTTGAATGTTGGGAGAGAGAAATATATGAGCTCGTGGGTCTCTTCAATCAGAAATG 9856  
QY 9856 CTTGAATGTTGGGAGAGAGAAATATATGAGCTCGTGGGTCTCTTCAATCAGAAATG 9915  
DB CTTGAATGTTGGGAGAGAGAAATATATGAGCTCGTGGGTCTCTTCAATCAGAAATG 9915  
QY 961 CCAATCAGGTCAAGTTTGTCTAATTTGTATGTGTGATGCTTCCCAAGGTATAT 1020  
DB CCAATCAGGTCAAGTTTGTCTAATTTGTATGTGTGATGCTTCCCAAGGTATAT 9916  
QY 9916 CCAATCAGGTCAAGTTTGTCTAATTTGTATGTGTGATGCTTCCCAAGGTATAT 9975  
DB CCAATCAGGTCAAGTTTGTCTAATTTGTATGTGTGATGCTTCCCAAGGTATAT 9975  
QY 1021 TAACTATATAGAGAGTTGTGACAAACAGATGATTAAGCTGGACACCGTGGCACAGC 1080  
DB TAACTATATAGAGAGTTGTGACAAACAGATGATTAAGCTGGACACCGTGGCACAGC 9976  
QY 9976 TAACTATATAGAGAGTTGTGACAAACAGATGATTAAGCTGGACACCGTGGCACAGC 10035  
DB TAACTATATAGAGAGTTGTGACAAACAGATGATTAAGCTGGACACCGTGGCACAGC 10035  
QY 1081 TCAATGTTCTACTGCTTGGAGGTTGAGAGAGGAGATGCTTGAACACAGTGTCAA 1140  
DB TCAATGTTCTACTGCTTGGAGGTTGAGAGAGGAGATGCTTGAACACAGTGTCAA 10036  
QY 10036 TCAATGTTCTACTGCTTGGAGGTTGAGAGAGGAGATGCTTGAACACAGTGTCAA 10095  
DB TCAATGTTCTACTGCTTGGAGGTTGAGAGAGGAGATGCTTGAACACAGTGTCAA 10095  
QY 1141 GGCACAGCTGGGCAACATPACAGATCTCTCTCAAAAAAAAAAAAAAAAAAGAA 1200  
DB GGCACAGCTGGGCAACATPACAGATCTCTCTCTCAAAAAAAAAAAAAAAAAAGAA 10096  
QY 10096 GGCACAGCTGGGCAACATPACAGATCTCTCTCTCAAAAAAAAAAAAAAAAAAGAA 10155  
DB GGCACAGCTGGGCAACATPACAGATCTCTCTCTCAAAAAAAAAAAAAAAAAAGAA 10155  
QY 1201 GAGAGAGGCGGGGCGTGTGCTACAGCTGTAAATCCAGACATTTGGAGGCGGAGCC 1260  
DB GAGAGAGGCGGGGCGTGTGCTACAGCTGTAAATCCAGACATTTGGAGGCGGAGCC 10156  
QY 10156 GAGAGAGGCGGGGCGTGTGCTACAGCTGTAAATCCAGACATTTGGAGGCGGAGCC 10215  
DB GAGAGAGGCGGGGCGTGTGCTACAGCTGTAAATCCAGACATTTGGAGGCGGAGCC 10215  
QY 1261 GGGCGGATACCTGTGTGTGAGAGATTTGAGACACAGCTGGCCAAATGGCAAAACCCGT 1320  
DB GGGCGGATACCTGTGTGTGAGAGATTTGAGAGATTTGAGACACAGCTGGCCAAATGGCAAAACCCGT 10216  
QY 10216 GGGCGGATACCTGTGTGTGAGAGATTTGAGAGATTTGAGACACAGCTGGCCAAATGGCAAAACCCGT 10275  
DB GGGCGGATACCTGTGTGTGAGAGATTTGAGAGATTTGAGACACAGCTGGCCAAATGGCAAAACCCGT 10275  
QY 1321 CTGTACTCAAAATGCAAAAAATTAGCAGGCGGTGAGAGGACCTGTATCCAGCTAC 1380  
DB CTGTACTCAAAATGCAAAAAATTAGCAGGCGGTGAGAGGACCTGTATCCAGCTAC 10276  
QY 10276 CTGTACTCAAAATGCAAAAAATTAGCAGGCGGTGAGAGGACCTGTATCCAGCTAC 10335  
DB CTGTACTCAAAATGCAAAAAATTAGCAGGCGGTGAGAGGACCTGTATCCAGCTAC 10335  
QY 1381 TTGGAGGCTGAGGAGAGAAATCCCTTGAACCCAGAGGTGAGGTTGACGTAAAGCTGA 1440  
DB TTGGAGGCTGAGGAGAGAAATCCCTTGAACCCAGAGGTGAGGTTGACGTAAAGCTGA 10336  
QY 10336 TTGGAGGCTGAGGAGAGAAATCCCTTGAACCCAGAGGTGAGGTTGACGTAAAGCTGA 10395  
DB TTGGAGGCTGAGGAGAGAAATCCCTTGAACCCAGAGGTGAGGTTGACGTAAAGCTGA 10395  
QY 1441 GATCGTCCGTTGCACTCCAGCTGGGCGACAGAGACAACTCTGTCTCAGAAAAAAA 1500  
DB GATCGTCCGTTGCACTCCAGCTGGGCGACAGAGACAACTCTGTCTCAGAAAAAAA 10396  
QY 10396 GATCGTCCGTTGCACTCCAGCTGGGCGACAGAGACAACTCTGTCTCAGAAAAAAA 10455  
DB GATCGTCCGTTGCACTCCAGCTGGGCGACAGAGACAACTCTGTCTCAGAAAAAAA 10455  
QY 1501 AAAAAAGAGAGAGAGAGAGAAAGAAACATATTTGGAGAGAGAGATGGGAGAGAT 1560  
DB AAAAAAGAGAGAGAGAGAGAAAGAAACATATTTGGAGAGAGAGATGGGAGAGAT 10456  
QY 10456 AAAAAAGAGAGAGAGAGAGAAAGAAACATATTTGGAGAGAGAGATGGGAGAGAT 10515  
DB AAAAAAGAGAGAGAGAGAGAAAGAAACATATTTGGAGAGAGAGATGGGAGAGAT 10515  
QY 1561 TGAAGAGAAATGTGTTTATCAACAAATGTAGAGCCAAATAGGAGATCCATTTTG 1620  
DB TGAAGAGAAATGTGTTTATCAACAAATGTAGAGCCAAATAGGAGATCCATTTTG 10516  
QY 10516 TGAAGAGAAATGTGTTTATCAACAAATGTAGAGCCAAATAGGAGATCCATTTTG 10575  
DB TGAAGAGAAATGTGTTTATCAACAAATGTAGAGCCAAATAGGAGATCCATTTTG 10575  
QY 1621 TCTCTTTGTTCTATTGTTTCCCTTAACAAGTGTCTTGAACAGTAGAATAATATTCGA 1680  
DB TCTCTTTGTTCTATTGTTTCCCTTAACAAGTGTCTTGAACAGTAGAATAATATTCGA 10576  
QY 10576 TCTCTTTGTTCTATTGTTTCCCTTAACAAGTGTCTTGAACAGTAGAATAATATTCGA 10635  
DB TCTCTTTGTTCTATTGTTTCCCTTAACAAGTGTCTTGAACAGTAGAATAATATTCGA 10635  
QY 1681 ATPACCAATCCCTGTGCGGTTATTAAGTACAGCACTTTGCAATGAATGAGCAATCC 1740  
DB ATPACCAATCCCTGTGCGGTTATTAAGTACAGCACTTTGCAATGAATGAGCAATCC 10636  
QY 10636 ATPACCAATCCCTGTGCGGTTATTAAGTACAGCACTTTGCAATGAATGAGCAATCC 10695  
DB ATPACCAATCCCTGTGCGGTTATTAAGTACAGCACTTTGCAATGAATGAGCAATCC 10695  
QY 1741 ACAGAGAACTGAGAGACAACTGTCTATTTTAACTGTTTGAATGATGTTTGA 1800  
DB ACAGAGAACTGAGAGACAACTGTCTATTTTAACTGTTTGAATGATGTTTGA 10696  
QY 10696 ACAGAGAACTGAGAGACAACTGTCTATTTTAACTGTTTGAATGATGTTTGA 10755  
DB ACAGAGAACTGAGAGACAACTGTCTATTTTAACTGTTTGAATGATGTTTGA 10755



PR	14-ANG-2000	2000US-0225267P
PR	14-ANG-2000	2000US-0225266P
PR	14-ANG-2000	2000US-0225270P
PR	14-ANG-2000	2000US-022547P
PR	14-ANG-2000	2000US-0225757P
PR	14-ANG-2000	2000US-0225757P
PR	14-ANG-2000	2000US-0225759P
PR	18-ANG-2000	2000US-0225273P
PR	22-ANG-2000	2000US-0225681P
PR	22-ANG-2000	2000US-0225686P
PR	22-ANG-2000	2000US-0227118P
PR	22-ANG-2000	2000US-0227109P
PR	30-ANG-2000	2000US-0225924P
PR	01-SEP-2000	2000US-0228287P
PR	01-SEP-2000	2000US-0229343P
PR	01-SEP-2000	2000US-0229344P
PR	01-SEP-2000	2000US-0229345P
PR	05-SEP-2000	2000US-0229509P
PR	05-SEP-2000	2000US-0229513P
PR	06-SEP-2000	2000US-0230437P
PR	06-SEP-2000	2000US-0230438P
PR	06-SEP-2000	2000US-0231242P
PR	08-SEP-2000	2000US-0231243P
PR	08-SEP-2000	2000US-0231244P
PR	08-SEP-2000	2000US-0231413P
PR	08-SEP-2000	2000US-0231414P
PR	08-SEP-2000	2000US-0232080P
PR	08-SEP-2000	2000US-0232081P
PR	12-SEP-2000	2000US-0231968P
PR	14-SEP-2000	2000US-0232397P
PR	14-SEP-2000	2000US-0232398P
PR	14-SEP-2000	2000US-0232399P
PR	14-SEP-2000	2000US-0232400P
PR	14-SEP-2000	2000US-0232401P
PR	14-SEP-2000	2000US-0233063P
PR	14-SEP-2000	2000US-0233064P
PR	14-SEP-2000	2000US-0233065P
PR	21-SEP-2000	2000US-0234233P
PR	21-SEP-2000	2000US-0234374P
PR	25-SEP-2000	2000US-0234997P
PR	25-SEP-2000	2000US-0234998P
PR	26-SEP-2000	2000US-0235484P
PR	27-SEP-2000	2000US-0235834P
PR	27-SEP-2000	2000US-0235836P
PR	29-SEP-2000	2000US-0236327P
PR	29-SEP-2000	2000US-0236377P
PR	29-SEP-2000	2000US-0236368P
PR	29-SEP-2000	2000US-0236369P
PR	29-SEP-2000	2000US-0236370P
PR	02-OCT-2000	2000US-0237032P
PR	02-OCT-2000	2000US-0237037P
PR	02-OCT-2000	2000US-0237038P
PR	02-OCT-2000	2000US-0237039P
PR	13-OCT-2000	2000US-0237934P
PR	13-OCT-2000	2000US-0237935P
PR	13-OCT-2000	2000US-0239933P
PR	20-OCT-2000	2000US-0240960P
PR	20-OCT-2000	2000US-0241212P
PR	20-OCT-2000	2000US-0241785P
PR	20-OCT-2000	2000US-0241786P
PR	20-OCT-2000	2000US-0241787P
PR	20-OCT-2000	2000US-0241808P
PR	20-OCT-2000	2000US-0241809P
PR	20-OCT-2000	2000US-0241866P
PR	01-NOV-2000	2000US-0244617P
PR	08-NOV-2000	2000US-0246474P
PR	08-NOV-2000	2000US-0246475P
PR	08-NOV-2000	2000US-0246476P
PR	08-NOV-2000	2000US-0246477P
PR	08-NOV-2000	2000US-0246478P
PR	08-NOV-2000	2000US-0246523P
PR	08-NOV-2000	2000US-0246524P
PR	08-NOV-2000	2000US-0246525P



08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246529P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250319P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 06-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251892P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;  
WPI; 2001-483426/52.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
useful for preventing, diagnosing and/or treating cancers and metastasis,

Disclosure; SEQ ID NO 35900; 3071pp + Sequence Listing; English.

AAX54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic  
activity, and can be used in gene therapy and vaccine production. (I)  
proteins and polynucleotides may be used in the prevention, diagnosis and  
treatment of diseases associated with inappropriate (I) expression. For  
example, they may be used to treat disorders associated with decreased  
expression by rectifying mutations or deletions in a patient's genome  
that affect the activity of (I) by expressing inactive proteins or to  
supplement the patients own production of (I). Additionally, (I)  
polynucleotides may be used to produce the secreted (I), by inserting the  
nucleic acids into a host cell and culturing the cell to express the  
protein. (I) proteins and polynucleotides may be used to prevent,  
diagnose and treat immune/hematopoietic-related diseases, especially  
cancers and cancer metastases or hematopoietic-derived cells. AAK64703  
to AAK87694 represent human immune/haematopoietic antigen genomic  
sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
represent sequences used in the exemplification of the present invention

Sequence 1665 BP; 497 A; 343 C; 365 G; 460 T; 0 U; 0 Other;

Query Match 68.5%; Score 1488; DB 4; Length 1665;

	Best Local Similarity	100.0%	Pred. No. 6e-268	Mismatches	0	Indels	0	Gaps	0
QY	1	CCCCGTATGAGCTGATGACTTACGAGACCGATTGCTACTTTGGGACTTTGATAGCA	60						
Db	178	CCCCGTATGAGCTGATGACTTACGAGACCGATTGCTACTTTGGGACTTTGATAGCA	237						
QY	61	CTTCCAGAAATGTGCACACAGATGAAATATCTCTGCTGACAGACAGTGAATAAACAGT	120						
Db	238	CTTCCAGAAATGTGCACACAGATGAAATATCTCTGCTGACAGACAGTGAATAAACAGT	297						
QY	121	CCCTCAAGTCTCTCTGTTTTTATCTTCAACTCTCACTTCTTTAGATTACAGAAAAA	180						
Db	238	CCCTCAAGTCTCTCTGTTTTTATCTTCAACTCTCACTTCTTTAGATTACAGAAAAA	357						
QY	181	ATATTATATACGACTCTTTAAAAAGATCTATGCTTGAAAAATAGAGAAACACAGT	240						
Db	358	ATATTATATACGACTCTTTAAAAAGATCTATGCTTGAAAAATAGAGAAACACAGT	417						
QY	241	CTGGCCAGGAGACGTGCTGCAATTTGGTGAAGTTTGAATGCAACATGTCCTCTGCGGA	300						
Db	418	CTGGCCAGGAGACGTGCTGCAATTTGGTGAAGTTTGAATGCAACATGTCCTCTGCGGA	477						
QY	301	ATPACAGAACTGACAGACCTGGGAGCATCTPAAAGTGTCAACGTTTTCTPAGACTTTTA	360						
Db	478	ATPACAGAACTGACAGACCTGGGAGCATCTPAAAGTGTCAACGTTTTCTPAGACTTTTA	537						
QY	361	GGTAGATGAGACGAGAAAGTATCTCTPAAAAGCATGTTAGAGAGATCAATGTTTTTA	420						
Db	538	GGTAGATGAGACGAGAAAGTATCTCTPAAAAGCATGTTAGAGAGATCAATGTTTTTA	597						
QY	421	TATCAACATCCTTATATTTGATTCATTGATTAACAGTGGTATGATATATTTT	480						
Db	538	TATCAACATCCTTATATTTGATTCATTGATTAACAGTGGTATGATATATTTT	657						
QY	481	TCTATCTTCCCTGAGCTTCTTCAAGTATACAACTCTTCATCAGGCGCATGA	540						
Db	658	TCTATCTTCCCTGAGCTTCTTCAAGTATACAACTCTTCATCAGGCGCATGA	717						
QY	541	TCTATAGACCTCCTAATGAGAGATCTGGGTGATGTGACCCCAACATCTCTCCAA	600						
Db	718	TCTATAGACCTCCTAATGAGAGATCTGGGTGATGTGACCCCAACATCTCTCCAA	777						
QY	601	GCATTAAATCCATCATGCGCTGATGTTTTAATCAGACAGACATGTTTTATGTTG	660						
Db	778	GCATTAAATCCATCATGCGCTGATGTTTTAATCAGACAGACATGTTTTATGTTG	837						
QY	661	TACAAAABAAGTTGTTATGGGTGGGAGTGAAGGATAGACATGACATGTCACCTTCAA	720						
Db	838	TACAAAABAAGTTGTTATGGGTGGGAGTGAAGGATAGACATGACATGTCACCTTCAA	897						
QY	721	GCTACTTAAATAAAGATCTTAAATGCGGACGAGACCTGTGAACAAGACCCCTAATA	780						
Db	898	GCTACTTAAATAAAGATCTTAAATGCGGACGAGACCTGTGAACAAGACCCCTAATA	957						
QY	781	TGGGTGATGTCTGAAGTGAACAATCTTCTGAAAACGCAAACTCTTTTAAAGAACTCCT	840						
Db	958	TGGGTGATGTCTGAAGTGAACAATCTTCTGAAAACGCAAACTCTTTTAAAGAACTCCT	1011						
QY	841	AATTATGAAAACCCCAAACTTTCACATATCAATAATTGCAAAACATTTGAAAGAAAGTTG	900						
Db	1018	AATTATGAAAACCCCAAACTTTCACATATCAATAATTGCAAAACATTTGAAAGAAAGTTG	1077						
QY	901	CTTGAATGTGGGAGAGAAATCTATTTGGCTCTCGTGGTCTCTTCACTCAGAAATG	960						
Db	1078	CTTGAATGTGGGAGAGAAATCTATTTGGCTCTCGTGGTCTCTTCACTCAGAAATG	1133						
QY	961	CCAAATCAGCTCAAGTTTGCTACATTTTGTATGTGTGATGCTTCTCCAAAGGTATAT	1020						
Db	1138	CCAAATCAGCTCAAGTTTGCTACATTTTGTATGTGTGATGCTTCTCCAAAGGTATAT	1197						
QY	1021	TAACTATTAAGAGCTTGTGCAAAACAGAAATGATAAAGCTGCGAACCGTGGCACAGGC	1088						



Db 1198 TAACATATATAGAGAGTTGTGACAAACAGAAATGATAAGAGTCGCAACCTGGACACGC 1257  
 Qy 1081 TCATAGTTTCTAGCTGCTGGAGAGTTGAGAGGAGAGTGGCTTGAACACAGGTGTTCAA 1140  
 Db 1258 TCATAGTTTCTAGCTGCTGGAGAGTTGAGAGGAGAGTGGCTTGAACACAGGTGTTCAA 1317  
 Qy 1141 GGCCAGCCTGGGCAACATACAAAGTCTGCTCTCAAAAAAAGAAAAAAGAAAAA 1200  
 Db 1318 GGCCAGCCTGGGCAACATACAAAGTCTGCTCTCAAAAAAAGAAAAAAGAAAAA 1377  
 Qy 1201 GAGAGAGGAGGAGGAGGAGTGTGATCAAGCTGTATCCACAGCACTTTGGAGGCGGAGCC 1260  
 Db 1378 GAGAGAGGAGGAGGAGGAGTGTGATCAAGCTGTATCCACAGCACTTTGGAGGCGGAGCC 1437  
 Qy 1261 GGGCGGATCACCCTGTGTGATGAGAGTTTGAACACAGCTGGCCAAACATGGCAAAACCCCGT 1320  
 Db 1438 GGGCGGATCACCCTGTGTGATGAGAGTTTGAACACAGCTGGCCAAACATGGCAAAACCCCGT 1497  
 Qy 1321 CTGTACTCAAAATGCAAAATTAATGACAGGCGCTGTGAGAGGACACCTGTAAATCCAGGCTAC 1380  
 Db 1498 CTGTACTCAAAATGCAAAATTAATGACAGGCGCTGTGAGAGGACACCTGTAAATCCAGGCTAC 1557  
 Qy 1391 TTGGAGGCTGAGGAGGAGATGCTTGAACCCAGAGAGTGAAGTGAAGTGA 1440  
 Db 1558 TTGGAGGCTGAGGAGGAGATGCTTGAACCCAGAGAGTGAAGTGAAGTGA 1617  
 Qy 1441 GATCGTGCCTGTGCACTCCAGCCTGGGCGACAGAGCAAGCACTGTGTC 1488  
 Db 1618 GATCGTGCCTGTGCACTCCAGCCTGGGCGACAGAGCAAGCACTGTGTC 1665

RESULT 9  
 ABX05468  
 ID ABX05468 standard; cDNA; 1543 BP.

17-JAN-2003 (first entry)  
 Human novel polynucleotide #483.  
 Human; gene; ss; genetic disorder; gene mapping; medical imaging; cancer;  
 neurodegenerative disorder; lymphoid cell disorder; osteoporosis;  
 Parkinson's disease; Alzheimer's disease; bone degenerative disorder;  
 osteoarthritis; periodontal disease; liver fibrosis; viral infection;  
 fungal infection; bacterial infection; autoimmune disease; diabetes;  
 atopic dermatitis.  
 Homo sapiens.  
 WO200274961-A1.  
 26-SEP-2002.  
 14-MAR-2002; 2002WO-US005109.  
 15-MAR-2001; 2001US-00810173.  
 (HYSE-) HISEQ INC.  
 Tang YT, Zhou P, Goodrich R, Asundi V, Zhang J, Zhao QA, Ren F,  
 Pi Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M,  
 Pi Wehrman T, Wang J, Wang D, Dmanac RT;  
 MPI; 2003-040556/03.  
 P-PSDB; ABU00390.  
 New isolated polypeptides and polynucleotides, useful for preventing,  
 treating or ameliorating medical conditions, such as cancer,  
 neurodegenerative disorders, lymphoid cell disorders, bone degenerative  
 disorders, and infections.  
 Claim 1; SEQ ID NO 483; 235pp; English.

XX The invention relates to human polynucleotides and the polypeptides they  
 CC encode. The polynucleotides and polypeptides are useful in diagnostics,  
 CC forensics, gene mapping, medical imaging, identification of mutations  
 CC responsible for genetic disorders or other traits, assessing biodiversity  
 CC and producing many other types of data and products dependent on DNA and  
 CC amino acid sequences. They are also useful for preventing, treating or  
 CC ameliorating medical conditions, such as cancer, neurodegenerative  
 CC disorders (e.g. Parkinson's disease, Alzheimer's disease), lymphoid cell  
 CC disorders, osteoporosis, osteoarthritis, bone degenerative disorders,  
 CC periodontal disease, liver fibrosis, infections (e.g. viral, fungal or  
 CC bacterial) or autoimmune diseases (e.g. diabetes, atopic dermatitis).  
 CC Sequences ABX04986-ABX05511 represent human polynucleotides of the  
 CC invention. Note: The sequence data for this patent is not represented in  
 CC the printed specification but is based on sequence information supplied  
 CC by the European Patent Office

SQ Sequence 1543 BP; 428 A; 334 C; 352 G; 429 T; 0 U; 0 Other;

Query Match 42.5%; Score 923; DB 7; Length 1543;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-162; Mismatches 0; Gaps 0;

Matches 923; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 CCCCTGTATGAGTTGATGACTTACGAGACGCACTTCTGATCTTGGACCTTGTATGACA 60  
 Db 621 CCCCTGTATGAGTTGATGACTTACGAGACGCACTTCTGATCTTGGACCTTGTATGACA 680  
 Qy 61 CTTCCAGAGATGTACACACAGATGAATATCTCTGCTGAAGACATGTGATATAAAACAGT 120  
 Db 681 CTTCCAGAGATGTACACACAGATGAATATCTCTGCTGAAGACATGTGATATAAAACAGT 740  
 Qy 121 CCTTCAAGCTTCTCTGTTTATTTCTTCACTCACTTCTGAGTTTACAGAAAAA 180  
 Db 741 CCTTCAAGCTTCTCTGTTTATTTCTTCACTCACTTCTGAGTTTACAGAAAAA 800  
 Qy 181 ATATTATATACAGACTCTTAAAGATCTATGCTTGAATAATAGAGAACACAGGT 240  
 Db 801 ATATTATATACAGACTCTTAAAGATCTATGCTTGAATAATAGAGAACACAGGT 860  
 Qy 241 CTGGCAGAGGACGCTGTGCAATTTGTGCAAGTTTGAATCAACATTTCCCTACTGGGA 300  
 Db 861 CTGGCAGAGGACGCTGTGCAATTTGTGCAAGTTTGAATCAACATTTCCCTACTGGGA 920  
 Qy 301 ATACAGAGCTGAGAGACCTGGAGAGATCTTAAAGTGAAGTTTCTATGACTTTA 360  
 Db 921 ATACAGAGCTGAGAGACCTGGAGAGATCTTAAAGTGAAGTTTCTATGACTTTA 980  
 Qy 361 GGTAGATGAGAGCAGAGAGTATCTTAAAGATGCTGAGAGATCAAAATGTTTTA 420  
 Db 981 GGTAGATGAGAGCAGAGAGTATCTTAAAGATGCTGAGAGATCAAAATGTTTTA 1040  
 Qy 421 TATCAACATCCTTATTTATTTGATTCATTTGATTAACAGTGTGTATGATATTT 480  
 Db 1041 TATCAACATCCTTATTTATTTGATTCATTTGATTAACAGTGTGTATGATATTT 1100  
 Qy 481 TCTATCTTTTCCCTGACCTTACTTTCAAGTAAACAAACCTTCCATCGGCCATGA 540  
 Db 1101 TCTATCTTTTCCCTGACCTTACTTTCAAGTAAACAAACCTTCCATCGGCCATGA 1160  
 Qy 541 TCTATGAGACCTGCTATGAGATATCTGGGATGTGTGACCCCAACATCTCCAAA 600  
 Db 1161 TCTATGAGACCTGCTATGAGATATCTGGGATGTGTGACCCCAACATCTCCAAA 1220  
 Qy 601 GCATTATATCAATCATGCGCTGTATGTTTATACAGCAAGACATGTTTATGTTTG 660  
 Db 1221 GCATTATATCAATCATGCGCTGTATGTTTATACAGCAAGACATGTTTATGTTTG 1280  
 Qy 661 TACAAAAGAAATTTTATGAGTGGAGATGAGATATGACCATGATGATCTTCA 720  
 Db 1281 TACAAAAGAAATTTTATGAGTGGAGATGAGATATGACCATGATGATCTTCA 1340  
 Qy 721 GCTACTTATATTAAGATCTTAAATGAGAGAGCTGTGAACAGACACCTTATA 780



Db 1341 GCTACTTTAATAAGATCTTAAATGGCAGAGGACTGTGACACAGACCTTAATTA 1400  
QY 781 TGGGTGATGCTGAGTAGTGAATCTTGTGAAAAGCAACTTTTAAGAACTCCCT 840  
Db 1401 TGGGTGATGCTGAGTAGTGAATCTTGTGAAAAGCAACTTTTAAGAACTCCCT 1460  
QY 841 AATTGGAACACCCCAACTTCATATCATATTAATTGCAACAACTGGAGAAGTTG 900  
Db 1461 AATTGGAACACCCCAACTTCATATCATATTAATTGCAACAACTGGAGAAGTTG 1520  
QY 901 CTTGAATGTTGGGAGAGGAAAA 923  
Db 1521 CTTGAATGTTGGGAGAGGAAAA 1543  
  
RESULT 10  
AAK81089  
ID AAK81089 standard; DNA; 574 BP.  
XX AAK81089;  
AC AAK81089;  
XX  
XX 07-NOV-2001 (first entry)  
XX  
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:35901.  
XX  
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytostatic; gene therapy; vaccine; metastasis; ds.  
XX  
OS Homo sapiens.  
XX  
PN M0200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 17-JAN-2001; 2001MO-US001354.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214866P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218990P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225113P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225577P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226688P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.

PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232337P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236337P.  
PR 29-SEP-2000; 2000US-0236337P.  
PR 29-SEP-2000; 2000US-0236338P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241211P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246533P.  
PR 08-NOV-2000; 2000US-0246534P.  
PR 08-NOV-2000; 2000US-0246535P.  
PR 08-NOV-2000; 2000US-0246536P.  
PR 08-NOV-2000; 2000US-0246572P.  
PR 08-NOV-2000; 2000US-0246588P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.



PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249219P.  
PR 17-NOV-2000; 2000US-0249220P.  
PR 17-NOV-2000; 2000US-0249221P.  
PR 17-NOV-2000; 2000US-0249222P.  
PR 17-NOV-2000; 2000US-0249223P.  
PR 17-NOV-2000; 2000US-0249224P.  
PR 17-NOV-2000; 2000US-0249225P.  
PR 17-NOV-2000; 2000US-0249226P.  
PR 17-NOV-2000; 2000US-0249227P.  
PR 17-NOV-2000; 2000US-0249228P.  
PR 17-NOV-2000; 2000US-0249229P.  
PR 17-NOV-2000; 2000US-0249230P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251888P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251899P.  
PR 08-DEC-2000; 2000US-0251909P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-483426/52.  
DR  
XX  
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides.  
PT Useful for preventing, diagnosing and/or treating cancers and metastasis.  
PS  
XX  
XX Disclosure; SEQ ID NO 35901; 3071bp + Sequence Listing; English.  
PS  
XX  
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
CC amino acid sequences given in AAK62170 to AAK61921. (I) have cytosolic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patient's own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting the  
CC nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/hematopoietic-related diseases, especially  
CC cancers and cancer metastases of hematopoietic-derived cells. AAK4703  
CC to AAK87694 represent human immune/hematopoietic antigen genomic  
CC sequences from the present invention. AAK54952 to AAK54950 and AAK62169  
CC represent sequences used in the exemplification of the present invention  
CC  
XX  
SQ Sequence 574 BP; 201 A; 80 C; 85 G; 208 T; 0 U; 0 Other;

Query Match 26.4%; Score 574; DB 4; Length 574;  
Best Local Similarity 100.0%; Pred. No. 1e-97;  
Matches 574; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1589 AATGTAGAGAGCAATTAAGGATCCCTATTGCTCTTTGGTGTCTATTGTCCTTAC 1648  
DB 1 AATGTAGAGAGCAATTAAGGATCCCTATTGCTCTTTGGTGTCTATTGTCCTTAC 60  
QY 1649 AACTGCTTTGACAGTGAGAAAATATTCAGAAATACATATCCCTGCGCTTATACC 1708  
DB 61 AACTGCTTTGACAGTGAGAAAATATTCAGAAATACATATCCCTGCGCTTATACC 120  
QY 1709 TAGCAACCTTGGAAATGAGATAGAGATCCACAGAGAAACTTGAATGCACAACTGCT 1768  
DB 121 TAGCAACCTTGGAAATGAGATAGAGATCCACAGAGAAACTTGAATGCACAACTGCT 180

QY 1769 TATTTTAACCTTATTTGATACAGTTTGTGAAAAGATTAAAAATTGTAATTCATGTATT 1828  
DB 181 TATTTTAACCTTATTTGATACAGTTTGTGAAAAGATTAAAAATTGTAATTCATGTATT 240  
QY 1829 CATTATATTTATTTATTTATTTTGGCTGATGATTTTATTTATTAACATGATTCCTTTCT 1888  
DB 241 CATTATATTTATTTATTTATTTTGGCTGATGATTTTATTTATTAACATGATTCCTTTCT 300  
QY 1889 GATATATTTGAAATGAGCTGCAAGCTTATTAATTTATTAACCTTAGAATGATCTAAT 1948  
DB 301 GATATATTTGAAATGAGCTGCAAGCTTATTAATTTATTAACCTTAGAATGATCTAAT 360  
QY 1949 AACACGATGATTTATTTGAAATTTGAGATGATGCTGCAAGCTTATTTGATTTCTGATTT 2008  
DB 361 AACACGATGATTTATTTGAAATTTGAGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420  
QY 2009 TTAGTAACTTTTATGACAGCAATTTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2068  
DB 421 TTAGTAACTTTTATGACAGCAATTTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480  
QY 2069 GATTAATATTTTGGAGCTGCTGCAATTTAATCCAAATTAATTAATTAATTAATTAATTAAT 2128  
DB 481 GATTAATATTTTGGAGCTGCTGCAATTTAATCCAAATTAATTAATTAATTAATTAATTAAT 540  
QY 2129 TATATGAGTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2162  
DB 541 TATATGAGTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 574  
RESULT 11  
AAK61819  
ID AAK61819 standard; cDNA; 429 BP.  
XX  
AC AAK61819;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human immune/hematopoietic antigen encoding cDNA SEQ ID NO:6879.  
XX  
KW Human; immune; hematopoietic; immune/hematopoietic antigen; cancer;  
KW cytosolic; gene therapy; vaccine; metastasis; ss.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
XX  
OS Homo sapiens.  
XX  
PN MO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
FF 17-JAN-2001; 2001MO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184



PR	14-AUG-2000	2000US-02255666
PR	14-AUG-2000	2000US-02255676
PR	14-AUG-2000	2000US-02255686
PR	14-AUG-2000	2000US-02255700
PR	14-AUG-2000	2000US-02255710
PR	14-AUG-2000	2000US-02255747
PR	14-AUG-2000	2000US-02255757
PR	14-AUG-2000	2000US-02255758
PR	14-AUG-2000	2000US-02255759
PR	18-AUG-2000	2000US-02265799
PR	18-AUG-2000	2000US-02266810
PR	22-AUG-2000	2000US-02266868
PR	22-AUG-2000	2000US-02271822
PR	23-AUG-2000	2000US-02270099
PR	30-AUG-2000	2000US-02286249
PR	01-SEP-2000	2000US-02292877
PR	01-SEP-2000	2000US-02293433
PR	01-SEP-2000	2000US-02293443
PR	01-SEP-2000	2000US-02293457
PR	05-SEP-2000	2000US-02295099
PR	05-SEP-2000	2000US-02295613
PR	06-SEP-2000	2000US-02300378
PR	06-SEP-2000	2000US-02300437
PR	08-SEP-2000	2000US-02313422
PR	08-SEP-2000	2000US-02313433
PR	08-SEP-2000	2000US-02317444
PR	08-SEP-2000	2000US-02318143
PR	08-SEP-2000	2000US-02318149
PR	08-SEP-2000	2000US-02320808
PR	08-SEP-2000	2000US-02320818
PR	12-SEP-2000	2000US-02331968
PR	14-SEP-2000	2000US-02332977
PR	14-SEP-2000	2000US-02332987
PR	14-SEP-2000	2000US-02332999
PR	14-SEP-2000	2000US-02334009
PR	14-SEP-2000	2000US-02334019
PR	14-SEP-2000	2000US-02336363
PR	14-SEP-2000	2000US-02336649
PR	14-SEP-2000	2000US-02336659
PR	21-SEP-2000	2000US-02342233
PR	21-SEP-2000	2000US-02342237
PR	21-SEP-2000	2000US-02342747
PR	25-SEP-2000	2000US-02349979
PR	25-SEP-2000	2000US-02349988
PR	25-SEP-2000	2000US-02349989
PR	26-SEP-2000	2000US-02355844
PR	27-SEP-2000	2000US-02355849
PR	27-SEP-2000	2000US-02355867
PR	29-SEP-2000	2000US-02363277
PR	29-SEP-2000	2000US-02363677
PR	29-SEP-2000	2000US-02365689
PR	29-SEP-2000	2000US-02365699
PR	29-SEP-2000	2000US-02365702
PR	29-SEP-2000	2000US-02369335
PR	13-OCT-2000	2000US-02399377
PR	13-OCT-2000	2000US-02399379
PR	20-OCT-2000	2000US-02411212
PR	20-OCT-2000	2000US-02412866
PR	01-NOV-2000	2000US-02446177
PR	08-NOV-2000	2000US-02466747
PR	08-NOV-2000	2000US-02466749
PR	08-NOV-2000	2000US-02466756
PR	08-NOV-2000	2000US-02466777
PR	08-NOV-2000	2000US-02466787
PR	08-NOV-2000	2000US-02466799
PR	08-NOV-2000	2000US-02466823
PR	08-NOV-2000	2000US-02466829

[illegible]



Query Match 19.8%; Score 429; DB 4; Length 429;  
Best Local Similarity 100.0%; Pred. No. 1e-70;  
Matches 429; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	76	ACAAGACACCCATAAATAGGTTGATGCTGAATAGCAAACTCTTCGAAACGGAACT	823
Db	1	ACAGACACCCATATATATAGGTTGATGCTGAATAGCAAACTCTTCGAAACGGAACT	60
QY	824	CTTTTAAGGAAGTCCCTAATTTTGAACAACCCCAAACTTCAATATCATTAATAGCAAA	883
Db	61	CTTTTAAGGAAGTCCCTAATTTTGAACAACCCCAAACTTCAATATCATTAATAGCAAA	120
QY	884	CAATTGGAAGGAAGTTGCTTGAATGTTGGGAGAGGAAATCTAATTGGCTCTCGTGGGCT	943
Db	121	CAATTGGAAGGAAGTTGCTTGAATGTTGGGAGAGGAAATCTAATTGGCTCTCGTGGGCT	180
QY	944	TCCTCATCTCAGAAATGCCAATCAGGTCAAGTTTGCTACATTTTGTATGTGATGCTG	1003
Db	181	TCCTCATCTCAGAAATGCCAATCAGGTCAAGTTTGCTACATTTTGTATGTGATGCTG	240
QY	1004	TTCTCCCAAGGATATATTAATTAATAAGAGAGTTGTACAAAACGAATGATTAAGCTG	1063
Db	241	TTCTCCCAAGGATATATTAATTAATAAGAGAGTTGTGTACAAAACGAATGATTAAGCTG	300
QY	1064	CGAACCCTGACACAGCTCATATGTTCTATGCTGCTTGGAGGTTGAGGAGGAGAGATGGCT	1122
Db	301	CGAACCCTGACACAGCTCATATGTTCTATGCTGCTTGGAGGTTGAGGAGGAGAGATGGCT	360
QY	1124	TGAACACAGGTGTTTCAAGGCGAGCTGGGCAACATTAACAAGATCCTGCTCTCAAAAAA	1183
Db	361	TGAACACAGGTGTTTCAAGGCGAGCTGGGCAACATTAACAAGATCCTGCTCTCAAAAAA	420
QY	1184	AAAAAAAAA 1192	
Db	421	AAAAAAAAA 429	
RESULT 12			
ID	AA102789		
AA102789	standard; DNA; 13662 BP.		
AC	AA102789;		
XX	21-NOV-2001 (first entry)		
DE	Human reproductive system related antigen DNA SEQ ID NO: 5477.		
XX	Human; reproductive system related antigen; reproductive system disorder;		
KM	cancer; gene therapy; ds.		
XX	Homo sapiens.		
OS	WO200155320-A2.		
FN	02-AUG-2001.		
XX	17-JAN-2001; 2001WO-US001339.		
PF			
XX			
PR	31-JAN-2000; 2000US-0179065P.		
PR	04-FEB-2000; 2000US-0180628P.		
PR	24-FEB-2000; 2000US-0184664P.		
PR	02-MAR-2000; 2000US-0186350P.		
PR	16-MAR-2000; 2000US-0189874P.		
PR	17-MAR-2000; 2000US-0190076P.		
PR	18-APR-2000; 2000US-0198123P.		
PR	19-MAY-2000; 2000US-020515P.		
PR	07-JUN-2000; 2000US-0209467P.		
PR	28-JUN-2000; 2000US-0214886P.		
PR	30-JUN-2000; 2000US-0215135P.		
PR	07-JUL-2000; 2000US-0216647P.		
PR	07-JUL-2000; 2000US-0216880P.		
PR	11-JUL-2000; 2000US-0217487P.		



PR 01-NOV-2000; 2000US-0244617P.  
 PR 08-NOV-2000; 2000US-0246474P.  
 PR 08-NOV-2000; 2000US-0246475P.  
 PR 08-NOV-2000; 2000US-0246476P.  
 PR 08-NOV-2000; 2000US-0246477P.  
 PR 08-NOV-2000; 2000US-0246478P.  
 PR 08-NOV-2000; 2000US-0246523P.  
 PR 08-NOV-2000; 2000US-0246524P.  
 PR 08-NOV-2000; 2000US-0246525P.  
 PR 08-NOV-2000; 2000US-0246526P.  
 PR 08-NOV-2000; 2000US-0246527P.  
 PR 08-NOV-2000; 2000US-0246528P.  
 PR 08-NOV-2000; 2000US-0246532P.  
 PR 08-NOV-2000; 2000US-0246609P.  
 PR 08-NOV-2000; 2000US-0246610P.  
 PR 08-NOV-2000; 2000US-0246611P.  
 PR 08-NOV-2000; 2000US-0246613P.  
 PR 17-NOV-2000; 2000US-0249207P.  
 PR 17-NOV-2000; 2000US-0249208P.  
 PR 17-NOV-2000; 2000US-0249209P.  
 PR 17-NOV-2000; 2000US-0249210P.  
 PR 17-NOV-2000; 2000US-0249211P.  
 PR 17-NOV-2000; 2000US-0249212P.  
 PR 17-NOV-2000; 2000US-0249213P.  
 PR 17-NOV-2000; 2000US-0249214P.  
 PR 17-NOV-2000; 2000US-0249215P.  
 PR 17-NOV-2000; 2000US-0249216P.  
 PR 17-NOV-2000; 2000US-0249217P.  
 PR 17-NOV-2000; 2000US-0249218P.  
 PR 17-NOV-2000; 2000US-0249244P.  
 PR 17-NOV-2000; 2000US-0249245P.  
 PR 17-NOV-2000; 2000US-0249264P.  
 PR 17-NOV-2000; 2000US-0249265P.  
 PR 17-NOV-2000; 2000US-0249297P.  
 PR 17-NOV-2000; 2000US-0249298P.  
 PR 17-NOV-2000; 2000US-0249300P.  
 PR 01-DEC-2000; 2000US-0250160P.  
 PR 01-DEC-2000; 2000US-0250191P.  
 PR 05-DEC-2000; 2000US-0251030P.  
 PR 05-DEC-2000; 2000US-0251988P.  
 PR 05-DEC-2000; 2000US-0256719P.  
 PR 06-DEC-2000; 2000US-0251479P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 PR 08-DEC-2000; 2000US-0251989P.  
 PR 08-DEC-2000; 2000US-0251990P.  
 PR 11-DEC-2000; 2000US-0254097P.  
 PR 05-JAN-2001; 2001US-0259678P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX PI Rosen CA, Barash SC, Ruben SM;

XX DR WPI; 2001-465570/50.

XX PT Isolated nucleic acid molecule encoding a reproductive system antigen is used in preventing, treating or ameliorating a medical condition.

XX PS Disclosure; SEQ ID NO 5477; 1297pp + Sequence Listing; English.

XX CC The present invention provides the protein and coding sequences of a number of human reproductive system related antigens. These can be used in the prevention and treatment of reproductive system disorders, including cancer. The present sequence is a genomic sequence encoding a protein of the invention

XX SQ Sequence 13862 BP; 3188 A; 3560 C; 4465 G; 2649 T; 0 U; 0 Other;

Query Match 13.6%; Score 295.2; DB 4; Length 13862;  
 Best Local Similarity 74.6%; Pred. No. 1.5e-45;  
 Matches 397; Conservative 0; Mismatches 133; Indels 2; Gaps 2;

QY 1045 AAACAGATGATTAAGCTGCAACCGTGGACACGCTCATAGTCTAGCTGTTGGAGG 1104  
 DB 1508 AAACCTTAATAGCTGGCTGGGCTCTGGCTCAGGCCCTGATCCACGACTTTGGAGG 1567  
 QY 1105 TTGAGGAGGAGATGCTTTGAACAACAGTGTTCAGGCCAGCTGGCAACATTAACAG 1164  
 DB 1568 CTGAGGACGAGGATCATCTTAGAGTCAGAGTTTGAACACAGCTGGCCAACTAGCCAA 1627  
 QY 1165 ATCCGTCTCTCAAAAAAAAAAAAAAAAAAAGG-GAGAGGGCCGGCGTGGTGC 1223  
 DB 1628 ACCCATCTCTCAATAAATCAAAAGTTACCCAGGCTGGCCGGGATGTGGTGGC 1687  
 QY 1224 TCAGCCTGTATCCACGACTTTGGAGGCCGAGCCGGGGGATCACCTGTGTCAAGA 1283  
 DB 1688 TATGCTGTATCCACGACTTTGGAGGCCGAGTGGGGATCACTAGGTCAAGA 1747  
 QY 1284 GTTTGAGACGAGCTGGCCAAATGGCAAAACCCGCTGTACTCAAAATGCAAAATTA 1343  
 DB 1748 GTTCAAGACGAGCTGACCAACATAGCAAAACCGTCTATTAAAAATCAAAATTA 1807  
 QY 1344 GCCAGCGTGTAGGAGGACCTGTATCCAGCTACTTGGAGGCTGAGCGAGAGAT 1403  
 DB 1808 GCAGGCTTTATGTAGGAGCGCTTATATCCAGCTACTGGAGGCTGAGCGAGAGAT 1867  
 QY 1404 CGCTTGAAACCCAGAGGTGAGGTGGCTGAGTGAAGTGAATGTCGCTGACTCCAGCC 1463  
 DB 1868 CGTTTGAAGCCAGAGGGGGAGTTTGACCCGAGCCGAGATGTGCACTGCACTCCAGCC 1927  
 QY 1464 TGGGCGACAGAGCAGACTCTGTCTCAGAAAAAAGAGAGAGAGAGAGAA 1523  
 DB 1928 TGGGCGATTA-AGCAGACTCTGTCTCAGAAAAAAGAGAGAGAGAGAGAA 1986  
 QY 1524 AGAGAACATATTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1575  
 DB 1987 AAAAAAATTAGCCAGCGTGGTGGATGCACTGTAGTCTAGTACTTGGG 2038

RESULT 13  
 AAL07516  
 ID AAL07516 standard; DNA; 13862 BP.  
 XX  
 AC AAL07516;  
 XX  
 DT 21-NOV-2001 (first entry)  
 XX  
 DE Human reproductive system related antigen DNA SEQ ID NO: 10204.  
 XX  
 KW Human; reproductive system related antigen; reproductive system disorder;  
 KW cancer; gene therapy; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200155320-A2.  
 XX  
 PD 02-AUG-2001.  
 XX  
 PF 17-JAN-2001; 2001WO-US001339.  
 XX  
 PR 31-JAN-2000; 2000US-0179065P.  
 PR 04-FEB-2000; 2000US-0180628P.  
 PR 24-FEB-2000; 2000US-0184664P.  
 PR 02-MAR-2000; 2000US-0186350P.  
 PR 16-MAR-2000; 2000US-0189874P.  
 PR 17-MAR-2000; 2000US-0190076P.  
 PR 18-APR-2000; 2000US-0198123P.  
 PR 19-MAY-2000; 2000US-0205515P.  
 PR 07-JUN-2000; 2000US-0209467P.  
 PR 28-JUN-2000; 2000US-0214886P.  
 PR 30-JUN-2000; 2000US-0215135P.  
 PR 07-JUL-2000; 2000US-0216647P.  
 PR 07-JUL-2000; 2000US-0216800P.  
 PR 11-JUL-2000; 2000US-0217487P.  
 PR 11-JUL-2000; 2000US-0217496P.



```
PR 14-JUL-2000; 2000US-0218230P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226866P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 05-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 05-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0233397P.  
PR 14-SEP-2000; 2000US-0233398P.  
PR 14-SEP-2000; 2000US-0233399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234977P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0234984P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239335P.  
PR 13-OCT-2000; 2000US-0239337P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249246P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251889P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
  
(HUMA-) HUMAN GENOME SCI INC.  
XX  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-465570/50.  
XX  
XX Isolated nucleic acid molecule encoding a reproductive system antigen is  
PT used in preventing, treating or ameliorating a medical condition.  
XX  
XX  
XX Disclosure; SEQ ID NO 10204; 1297pp + Sequence Listing; English.  
XX  
XX  
XX The present invention provides the protein and coding sequences of a  
CC number of human reproductive system related antigens. These can be used  
CC in the prevention and treatment of reproductive system disorders,  
CC including cancer. The present sequence is a genomic sequence encoding a  
CC protein of the invention  
XX  
XX  
SQ Sequence 13862 BP; 3188 A; 3560 C; 4465 G; 2649 T; 0 U; 0 Other;  
  
Query Match 13.6%; Score 295.2; DB 4; Length 13862;  
Best Local Similarity 74.8%; Pred. No. 1.5e-45;  
Matches 397; Conservative 0; Mismatches 133; Indels 2; Gaps 2;  
  
QY 1045 AAACAGATGATTAAGCTGCAACCGTGGACACCGCTCATAGTCTAGCTGTGGAGG 1104
```



Db 1508 AAACCTATAGCTGGCTGGGCTCTGTGGCTCATGCTGTGATCCAGCACTTTGGAGG 1567  
Qy 1105 TTGGAGGAGGAGATGGCTTGAACACAGGTTCAGGCCAGCTGGGCACTAACAG 1164  
Db 1568 CTGAGGAGGAGATCACTTGAGGTCTAGAGTTGAACACAGCTGGCCCACTGCGAA 1627  
Qy 1165 ATCTGTCTCTCAAAAAAAAAAAAAAAAAAAGAAAG-CAGAGGGCGGGCGCTGTGGC 1223  
Db 1628 ACCCATCTCTACTTAAATAACAAAGTTAGCCAGGATGGGCGGCACTGTGGGTGC 1687  
Qy 1224 TCAGCCTGTATATCCAGCACTTTGGAGGCCGAGCCGGATCACTGTGGTCAAGA 1283  
Db 1688 TCATGCTGTATATCCAGCACTTTGGAGGCCGAGGCGGATCACTGTAGGTCAAGA 1747  
Qy 1284 GTTTGAGACCACTGTCGCAACATGSCAAAACCCCGTCTGTACTCAAAATGCAAAATTA 1343  
Db 1748 GTTCAGACCACTGTCGCAACATGSCAAAACCCCGTCTGTACTCAAAATGCAAAATTA 1807  
Qy 1344 GCCAGCGTGTGTAGCAGCACTGTATCCAGCTATTTGGAGGCTGAGGCGAGGAAT 1403  
Db 1808 GCCAGCGTGTATGTGTAGCCCTATATATCCAGCTACTGGAAGGCTGAGGCGAGGAAT 1867  
Qy 1404 CGCTTGAACCCAGGAGGTGAGGTTCAGTAAGCTGATCGTCCGTTGCACTCCAGGC 1463  
Db 1868 CGTTTGAAGCCAGGAGGCGAGGTTTGCACCGAGCCGATTTGCCACTGCACTCCAGGC 1927  
Qy 1464 TGGCGCAGAGGAGCACTCTGTCTCAGAAAAAAGAGAGAGAGAGAA 1523  
Db 1928 TGGCGCATTA-AGGAGACTGTCTCAGAAAAAAGAGAGAGAGAGAA 1986  
Qy 1524 AGAGAACATATTTTGGAGAGAGAGATGGGAAAGCATTGCAAGAAATTGTG 1575  
Db 1987 AGAAAAATTAGCCAGCGGTGTGGCATGACCTGTAGTCTTACTTCTTGGG 2038

RESULT 14  
ABA08208  
ID ABA08208 standard; DNA; 13862 BP.  
XX  
AC ABA08208;  
XX  
DT 11-JAN-2002 (first entry)  
XX  
DE Human ovarian and breast cancer associated polymucleotide SEQ ID NO 1003.  
XX  
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; anticancer;  
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; human; secreted protein; ds.  
XX  
OS Homo sapiens.  
XX  
PN WC00155325-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001MO-US001345.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198133P.  
PR 19-MAY-2000; 2000US-0205151P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226868P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235835P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 13-OCT-2000; 2000US-0239936P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.



PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX WPI, 2001-488786/53.  
XX  
XX  
XX New isolated ovarian and/or breast cancer related nucleic acids and  
PT polypeptides, useful for diagnosing, treating and/or preventing human  
PT diseases and disorders, particularly ovarian and/or breast cancer.  
XX  
XX  
XX Disclosure; SEQ ID NO 1003; 577bp + Sequence Listing; English.  
XX  
XX  
XX The invention relates to novel genes (ABA07454-ABA08224) and proteins  
CC (ABA07453-ABA10980) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in  
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,  
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune  
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic  
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,

Sequence 13862 BP; 3188 A; 3560 C; 4465 G; 2649 T; 0 U; 0 Other.

1045 AAGACGAATGATTAAGCTCGGAGCCG193CACACGCTCAATAG1CTACG19C11955859 11

1105 TTGAGCGAGGAGGA TTGGCTTGAACACAGGTGTTCAAGGCCAGCCTGGGCAACATACAG 116

Db 1568 CTGAGGCAGGAGTCACTTGAGGTCAAGAGTTGAGACCAAGCTGGCCACATGGCGAA 162

1165 ATCTGTCTCTCAAAAAAAAAAAAAAGAAACA-GAGAGGGCCGGGCGTGTGGC 122

[illegible]

Db 1688 TCATGCCCTGTAATCCACGACCTTTGGAGGCCGAGGTGGGCGATCACTGAGGTCAGGA 174

1284 GTTGAGACCAAGCTGGCCACATGCGAAAACCCCGTCTGTACTCAAAATGCAAAAATTAA 134

D5 I /48 GIIAHGACAGGC:GGACAAACAGCAGCGAGCCGCGGCTT 140

1808 GCCAGGCTTTATGTGTAAGCGGCTTATATCCAGCTACTCGAAGGCTGAGGACAGAGAAAT 1866

1404 CGCTTGAAACCCAGAGGTGGAGTTGCATTAAGCTGAGATCGTGCCTTGCACTCCAGCC 146

D5 1868 CGTTTGAAGCCAGAGCGGAGTTTCACCGAGCCGGGATGTGCTCACCACAGCACCCGCC

1928 TCCGCGATTA-AGCGAGCTGTCTCGAAAAAAGAAAAGAAAAGAAAA 198

1524 AGAGACAATATTTGGGAGAGAGGATGGGCAACATTGCAAGGAAATTGTG 1575

Db 1987 AGAAAAATTAGCCAGCGTGCGCATGCACCTGTAGTCCCTAGCTACTTGGG 2038

## RESULT 15

WP	Fragment Name	Begin	End	LOCUS	Accession
WP	Sequence split into 4 fragments			AC142/45	AC142/45

WP	ACF42745_1	100001	210000
MD	ACF42745_2	200001	310000

MP	ACF42/45_3	300001	354391
ID	ACF42/45 standard; DNA;	354391	BP.

AC ACF42745;  
XX

20-DEC-2000 14:40:00:00  
DA  
XX

XX Human; ALMS1; chromosome 2; 2p13; Alstrom disease; retinal dystrophy;  
KW

ophthalmological; antidiabetic; hepatotropic; nephrotropic; gene therapy

XX Homo sapiens.  
OS



Mon Mar 15 09:27:58 2004

us-09-966-880a-15.rng

Page 23

PX		FN	MOZ003034072-A2.	
XX		XD		
XX		ZD	24-APR-2003.	
XX		PF		
XX		FF	15-OCT-2002; 2002MCO-GB004658.	
XX		PR	15-OCT-2001; 2001GB-00024621.	
XX		PR	22-OCT-2001; 2001GB-00025318.	
XX		PR	07-JAN-2002; 2002GB-00000248.	
XX		PR	08-FEB-2002; 2002GB-00003039.	
XX		PR	08-FEB-2002; 2002GB-00003040.	
PA	(UYSO-) UNIV SOUTHAMPTON.			
PI	Wilson DI, Hearn T, Walker M;			
DR	WPI, 2003-393556/37.			
XX				
PT	Diagnosing the presence of, or susceptibility to, retinal dystrophy,			
PT	cardiomyopathy, endocrinopathy, diabetes, or Alstrom syndrome in an			
PT	individual, comprises detection or modulation of the ALMS1 protein or			
PT	gene region.			
PS	Disclosure; Page 42-101; 121pp; English.			
CC	The present invention describes a method of diagnosing the presence of,			
CC	or susceptibility to, retinal dystrophy, cardiomyopathy, endocrinopathy,			
CC	diabetes, or Alstrom syndrome in an individual. The method comprises			
CC	typing in a sample from the individual the ALMS1 protein or ALMS1 gene			
CC	region of the individual, or detecting aberrant ALMS1 activity. Human			
CC	ALMS1 is located to chromosome 2, more specifically to 2p13. ALMS1 has			
CC	ophthalmological, cardiac, antidiabetic, hepatotropic and nephrotropic			
CC	activities, and can be used in gene therapy. The method is useful for			
CC	diagnosing the presence of, or susceptibility to, retinal dystrophy,			
CC	cardiomyopathy, endocrinopathy (e.g. liver disease or renal impairment),			
CC	diabetes, or Alstrom syndrome in an individual. ALMS1 sequences can be			
CC	used in an agent that prevents or treats retinal dystrophy,			
CC	cardiomyopathy, endocrinopathy or diabetes is useful in manufacturing a			
CC	medicament for treating a patient who has been diagnosed as having or			
CC	being susceptible to retinal dystrophy, cardiomyopathy, endocrinopathy or			
CC	diabetes. ACP42632 to ACP42747 and ABR82113 to ABR82118 represent			
CC	sequences used in the exemplification of the present invention			
XX				
SQ	Sequence 354391 BP; 98637 A; 69688 C; 70813 G; 115253 T; 0 U; 0 Other;			
	Query Match 13.5%; Score 293.8; DB 7; Length 110000;			
	Best Local Similarity 74.3%; Pred. No. 3.6e-45;			
	Matches 384; Conservative 0; Mismatches 132; Indels 1; Gaps 1;			
OY	1000 ATGCTTCTCCCAAGATATATTAACTATATATAGAAGATTGTGCACAAAACAAGATATATAA	1059		
Dd	21564 ATTCTGTGCATTAATAACTTAAGAAAACTTCAGGAAGCATCATCTATAAAGTTAGATNG	21505		
OY	1060 GCTCGCAAACCGTGCCAACGGCTCATTAGTTTAGCTGTGGAGGTGGAGGAGAGAT	1119		
Dd	21504 GCTGCGGACGGGTGCTCATCTACCTGCCAGTCCCACACTTTAGAGGCGCAAGAGGAGAT	21445		
OY	1120 GGCTTGAACACAGGTGTCAAGGCGGCGGCGGCAACATPACAGAACCTGCTCTCAAA	1179		
Dd	21444 CGCTTGGAGCCAGAAATATATAGACACAGCCCTGGGCAACATPACATATCTCTTACAA	21385		
OY	1180 AAAAAAAAAAAAAAAAAAGAAAGAGAGAGGAGCGGCGGTGGCTCACCGCTGTATCCC	1239		
Dd	21384 AAAATTATTAATAATTAATAATTAACCGAGACCGGGCATGTGGCTCATGCCCTGTATCCC	21325		
OY	1240 AGCACTTTGGAGGCGGAGCCGGGCGGATCACTGTGTGTAGAGATTTAACCAACCAAGCTG	1299		
Dd	21324 AGCACTTTGGAGGCTAGAGGCGGATGTTCATGTGAGATCAGAAAGTTCCAGACCAAGCTG	21265		
OY	1300 GCCAACAATGGCAAAACCCTGCTACTCAAAATGCAAAATTAAGCCAGGCGGTGATAGA	1359		
Dd	21264 GCACAACATGTGAACCCCCGCTCTACTTAATAAATATAGCAAGCTTGTGGTGA	21205		

Qy	1360	GGGACCTGTATATCCCAAGTACTTTGGGAGGCTAGGAGGAGGAGTATGCTTGAACCGAGAG	1419
Db	21204	CATGCTCTGATATCCCAAGCTACTTTGAGAGGCTAGGTAGAGAAAT-GCTTGAACCTGGAG	21466
Qy	1420	GTCGAGGTTGCAGTAAAGTGAATCGTGCCTTTCGACTTCAGCTGGCGCAAGAGCAA	1479
Db	21145	GCGGAGGTTGCAGTGAGCCAAAGTCTTACATTGCATTCCAGCTGGGCAACAAGCGA	21086
Qy	1480	GACTCTGTCTCAGAAAAAAGAGAGAGA	1516
Db	21085	AATCTCATCTCAACAACAACAAAAATTTAACTGAGA	21049

Search completed: March 12, 2004, 18:08:15  
Job time : 874.74 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 12, 2004, 17:39:25 ; Search time 161.607 Seconds  
(without alignments)  
7458.540 Million cell updates/sec

Title: US-09-966-880A-15

Perfect score: 2172  
Sequence: 1 cccctgatgagtgatga.....agatgatgataacttg 2172

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Issued Patents NA:  
1: /cgn2\_6/ptodata/2/ina/5A COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCURS COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	280	15297	4 US-09-817-180-3	Sequence 3, Appl
2	280	15297	4 US-10-003-285-3	Sequence 3, Appl
3	274.2	12.9	US-09-078-284-7	Sequence 7, Appl
4	269.8	12.4	US-09-735-934A-3	Sequence 3, Appl
5	269.8	12.4	US-10-060-332-3	Sequence 3, Appl
6	269.2	12.4	US-09-798-036-10	Sequence 10, Appl
7	267.4	12.3	US-09-813-817-3	Sequence 3, Appl
8	267.4	12.3	US-09-978-187-3	Sequence 3, Appl
9	265.6	12.2	US-09-608-285A-8	Sequence 8, Appl
10	265.6	12.2	US-09-350-836B-8	Sequence 8, Appl
11	265.6	12.2	US-09-370-265-8	Sequence 8, Appl
12	265.6	12.2	US-09-557-800C-8	Sequence 8, Appl
13	265.6	12.2	US-09-370-625A-8	Sequence 8, Appl
14	265.6	12.2	US-09-608-285A-42	Sequence 42, Appl
15	265.6	12.2	US-09-557-800C-42	Sequence 42, Appl
16	265.6	12.2	US-09-608-285A-59	Sequence 59, Appl
17	262.2	12.1	US-09-810-671-3	Sequence 3, Appl
18	262.2	12.1	US-10-109-854-3	Sequence 3, Appl
19	260.4	12.0	US-09-788-172-18	Sequence 18, Appl
20	259.2	11.9	US-08-688-145-1	Sequence 1, Appl
21	258.6	11.9	US-09-735-934A-3	Sequence 3, Appl
22	258.6	11.9	US-10-060-332-3	Sequence 3, Appl
23	257.6	11.9	US-09-798-036-10	Sequence 10, Appl
24	257.2	11.8	US-08-618-100B-3	Sequence 3, Appl
25	256.8	11.8	US-07-914-281-5	Sequence 5, Appl
26	256.8	11.8	US-08-393-246-5	Sequence 5, Appl
27	256.8	11.8	US-08-525-058A-5	Sequence 5, Appl

C 28	256.8	11.8	8174	2	US-08-696-731-5	Sequence 5, Appl
C 29	256.8	11.8	8174	3	US-09-042-531-5	Sequence 5, Appl
C 30	256.8	11.8	8174	5	PCT-US91-00899-3	Sequence 3, Appl
C 31	256.8	11.8	29629	4	US-09-729-995-3	Sequence 3, Appl
C 32	256.8	11.8	29629	4	US-10-135-689-3	Sequence 3, Appl
C 33	254.8	11.7	118067	4	US-09-497-855A-32	Sequence 32, Appl
C 34	254.4	11.7	98844	4	US-09-791-211-10	Sequence 10, Appl
C 35	253.8	11.7	16063	4	US-09-801-053-3	Sequence 3, Appl
C 36	253.8	11.7	16063	4	US-10-020-121-3	Sequence 3, Appl
C 37	253.2	11.7	14581	4	US-08-520-373D-4	Sequence 4, Appl
C 38	253.2	11.7	22481	4	US-08-367-841A-43	Sequence 43, Appl
C 39	253.2	11.7	22481	5	PCT-US95-07201-43	Sequence 43, Appl
C 40	253.2	11.7	22484	4	US-09-875-223-2	Sequence 2, Appl
C 41	253.2	11.7	22484	4	US-09-875-114-2	Sequence 2, Appl
C 42	252.2	11.6	7210	2	US-08-257-963B-10	Sequence 10, Appl
C 43	252.2	11.6	7210	4	US-08-367-841A-10	Sequence 10, Appl
C 44	252.2	11.6	7210	5	PCT-US95-07201-10	Sequence 10, Appl
C 45	252.2	11.6	14581	4	US-08-520-373D-4	Sequence 4, Appl

## ALIGNMENTS

RESULT 1  
US-09-817-180-3  
Sequence 3, Application US/09817180  
Patent No. 6340584  
GENERAL INFORMATION:  
APPLICANT: GAN, Weiniu et al.  
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
FILE REFERENCE: CL001183  
CURRENT APPLICATION NUMBER: US/09/817,180  
CURRENT FILING DATE: 2001-03-27  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 3  
LENGTH: 15297  
TYPE: DNA  
ORGANISM: Human  
US-09-817-180-3

Query Match 12.9%; Score 280; DB 4; Length 15297;  
Best Local Similarity 74.9%; Pred. No. 9.3e-55;  
Matches 370; Conservative 0; Mismatches 110; Indels 14; Gaps 1;

QY	1019	ATTACTATATTAAGAGAGTTGACAAAACAGATGATTAAGTCCGAACGTCGACAC	1078
DB	11651	ACTCCCTTAATGCCAACCTTCCACACAGAGAAATAGATTAAGTCCGACGTTCTCAC	11710
QY	1079	GGCAGTACTTACGCTGCTGGAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1138
DB	11711	GCTGTCATCCAGACCTTGGAGGCTGAGTGGTGAATCACTTGAGCCACGAGATTC	11770
QY	1139	AAGGCAGCCTGAGCAATTAACAAGATCTGTCTCTCAAAAAAAAAAAAAAAAAAGA	1198
DB	11771	AAGATAGCTTGGACACACAGTGAATCTCTGTCACAAAAAATA-----	11825
QY	1199	AAGAGAGAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1258
DB	11826	-----GACTGGGACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	11876
QY	1259	CGGAGGATGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1318
DB	11877	GCAGGTGATCCCTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	11936
QY	1319	GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1378
DB	11937	ATCTCTACTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA	11996
QY	1379	ACTTGGAGGCTGAGGACAGAGATCGTTGAACCCAGAGAGAGAGAGAGAGAGAGAGAG	1438



Db 11997 ACTGGAGAGGCTGAGGTGGGAGAAATGCTTGAACCCAGAGGCGAGGCTGCAAGTGACC 12056  
QY 1439 GAGATGCGCGCTTGCACCTCCAGCGCTGGGCGACAAGACAGACTGTCTCAGAAAAAA 1498  
Db 12057 GAGATGTCGACCTGACCTCCAGCGCTGGGCGACAAGAGTGAATCCATCTCAAAAAAA 12116  
QY 1499 AAAAAAAAAAGAGA 1512  
Db 12117 CCAAAAAACAAAA 12130

## RESULT 2

US-10-003-295-3  
; Sequence 3, Application US/10003295  
; Patent No. 6686187  
; GENERAL INFORMATION:  
; APPLICANT: GAN, Weiniu et al.  
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
; TITLE OF INVENTION: THEREOF  
; FILE REFERENCE: CL001183DIV  
; CURRENT APPLICATION NUMBER: US/10/003,295  
; CURRENT FILING DATE: 2001-12-06  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 15297  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-003-295-3

Query Match 12.9%; Score 280; DB 4; Length 15297;  
Best Local Similarity 74.9%; Pred. No. 9.3e-55;  
Matches 370; Conservative 0; Mismatches 110; Indels 14; Gaps 1;

QY 1019 ATTACTATATPAGAGAGTGTGACAAACAGAAATGATTAAGCTGCGAACCGTGGCACAC 1078  
Db 11651 ACTCCCTTAATGCCAAGCTTCCACACAGGCAAGATTAAGAAATACCTGGCCAGTGTCTC 11710  
QY 1079 GCTCATAGTTCTAGCTGTGGAGGTTGAGAGGAGAGATGGCTTGAACAAGAGTGTTC 1138  
Db 11711 GCTGTATCCAGACACTTGGAGAGCTGAGCTGGGTGATCACTTGAGCCAGAGATTC 11770  
QY 1139 AAGCCAGCTTGGGCAACATTAACAGATCTGTCTCTCAAAAAAATAAAAAAAGA 1198  
Db 11771 AAGATCAGCTTGGACAAACAGTGAATCTCATCTGTCAAAAAATACAAAAATA----- 11825  
QY 1199 AAGAGAGAGGCGCGGCGGTGTGAGTCAAGCCTGTATCCAGACCTTGGAGGCGAG 1258  
Db 11826 -----GACTGGGACGCGGTGTGAGTCAAGCCTGTATCCAGACCTTGGAGGCGAG 11876  
QY 1259 CCGGCGGATCACTGTGTGTCAGAGTTGAGACCGAGCTGGCCAAATGGCAAAACCC 1318  
Db 11877 GAGGTGATCACTGTGTGTCAGAGTTGAGACCGAGCAACAGCAATGTGAACCC 11936  
QY 1319 GTCTGACTCAAAATGCAAAATTAACCGAGCGGTGTGACAGGCACTGTATCCAGCT 1378  
Db 11937 ATCTCTACTAAATAAATAAATAATTAAGCAGGATGTGGCACTGTATCCAGCT 11996  
QY 1379 ACTTGGAGGCTGAGGAGAGAAATCCGTTGAACCCAGAGGTGAGGTTGCAAGTAACT 1438  
Db 11997 ACTTGGAGGCTGAGGAGGAAATGCTTGAACCCAGAGGCGGAGCTGAGAGGCC 12056  
QY 1439 GAGATGTCGCTGACCTCCAGCGCTGGGCGACAAGACAGACTGTCTCAGAAAAAA 1498  
Db 12057 GAGATGTCGCTGACCTCCAGCGCTGGGCGACAAGAGTGAATCCATCTCAAAAAAA 12116  
QY 1499 AAAAAAAAAAGAGA 1512  
Db 12117 CCAAAAAACAAAA 12130

## RESULT 3

US-09-078-294-7/c  
; Sequence 7, Application US/09078294  
; Patent No. 6265211  
; GENERAL INFORMATION:  
; APPLICANT: Choo, Kong-Hong Andy  
; APPLICANT: Du Sart, Desiree  
; APPLICANT: Cancilla, Michael R.  
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE  
; FILE REFERENCE: Davies Col  
; CURRENT APPLICATION NUMBER: US/09/078,294  
; CURRENT FILING DATE: 1998-05-13  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 7  
; LENGTH: 11811  
; TYPE: DNA  
; ORGANISM: BAC-P2 contig 3  
US-09-078-294-7

Query Match 12.6%; Score 274.2; DB 3; Length 11811;  
Best Local Similarity 79.7%; Pred. No. 1.8e-53;  
Matches 350; Conservative 0; Mismatches 83; Indels 6; Gaps 2;

QY 1070 GTGGCACACGCTCATAGTTCTAGCTGTGGAGGTTGAGAGGAGATGCTTGAACA 1129  
Db 6663 GTGGCTCATGCTGTATATCCAGCACTTGGAGGCGAGGTGGAGATGATGATGAGCT 6604  
QY 1130 CAGGCTTAAAGCCAGCGCTGGGCAACATTAACAGATCTGTCTCAAAAAAATAA 1189  
Db 6603 CAGAGTTAAAGACCGAGCTGGGCAACAGTGAACCTCATCTTAACAAATACATTTA 6544  
QY 1190 AAAAAAAGAGAGAGAGGCGGCGGTGTGCTCAGCGCTGTATATCCAGCACTTTGG 1249  
Db 6543 AAGGTAG-----CTGGGGGCGAGGTGTGTGGCCACAGCTGTATCCAGCCTTTGG 6489  
QY 1250 GAGGCGGAGCGGCGGAGTACCTGTGTGACAGATTTGAGACAGCGCTGGCAACATGG 1309  
Db 6488 GAGGTCAAGGTGGGTGATCACTGAGGTCAAGAGTTGAGGCGAGCGCTGGCAACATGG 6429  
QY 1310 CAAACCCGCTGTATCTAATAATGC-AAAAATTAGCCAGGCTGTGACAGGCACTGT 1368  
Db 6428 TGAACCCCATCTCTAATAATAATAAATAAATTAGCCAGGCTGTGAGGCGGCTGT 6369  
QY 1369 AATCCAGTACTTGGAGGCTGAGGAGAGATCGCTTGAACCCAGAGGTTGAGGTT 1428  
Db 6368 AATCCAGTACTTGGAGGCTGAGGAGAGATCACTTGAACCCAGAGAGGTT 6309  
QY 1429 GCAGTAACTGAGATGTCGCTGGTGCACCTCCAGCTGGGCGACAAGAGCAACTGTGTC 1488  
Db 6308 GCAGTAACTGAGATGTCGCTGGTGCACCTCCAGCTGGGCGACAAGAGCAACTGTGTC 6249  
QY 1489 TCAGAAAAAATAAAAAA 1507  
Db 6248 AAAAAAAAAAAAAA 6230

## RESULT 4

US-09-735-934A-3/c  
; Sequence 3, Application US/09735934A  
; Patent No. 6372468  
; GENERAL INFORMATION:  
; APPLICANT: Li, Jiajin et al.  
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
; FILE REFERENCE: CL000851  
; CURRENT APPLICATION NUMBER: US/09/735,934A  
; CURRENT FILING DATE: 2000-12-14  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 43950  
; TYPE: DNA



ORGANISM: Homo sapiens  
US-09-735-934A-3

Query Match 12.4%; Score 269.8; DB 4; Length 43950;

Best Local Similarity 77.9%; Pred. No. 3.1e-52;

Matches 339; Conservative 0; Mismatches 92; Indels 4; Gaps 1;

QY 1063 GCGAACCTGTGGACACGCTCATATGTTCTGAGTGTGGAGGTTAGAGAGGAGATGGC 1122  
DB 17761 GGGCACAAGTGGCACAACCTGTATCTCCGAACTTTGGAGGCTAGGTGGAGATGGC 17702  
QY 1123 TTGAACACAGGTGTTCAGGCGCAGCTGGCAACATTAACAGATCTGTCTCAAAAAA 1182  
DB 17701 TTGAGGCGAGAGTTTCAAGACCACTGGGCAACATAGACAGACCCCTATCTTAAAAA 17642  
QY 1183 AAAAAAAAAAAAAA---AAGAGAGAGGCGCGGCGTGTGTCTACGCTGTATCC 1238  
DB 17641 AAAAAAAAAAAAAAAGAGAGAAATATAGTCAAGCGCGGTGTGTCTATGCTGTAAATCC 17582  
QY 1239 CAGCACTTTGGAGGCGCGGCGGAGTCACTGTGTCAAGATTTGAGACCAAGCT 1238  
DB 17581 CAGCACTTTGGAGGCGCGGCGGAGTCACTGTGTCAAGATTTGAGACCAAGCT 17522  
QY 1299 GCGCAACATGGCAAAACCCCTCTGTACTCAAAATGCAAAATTTAGCCAGCGTGTAGC 1358  
DB 17521 GACCAACATGGCAAAACCCCTCTGTACTCAAAATGCAAAATTTAGCTGGGCTGTAGC 17462  
QY 1359 AGGCACTGTATCCAGCTTCTGTAGGAGCTGGAGGAGAGATCGTTGAACCAAGCA 1418  
DB 17461 GGGTGTGTGTATCCAGCTTCTGTAGGAGCTGGAGGAGAGATCGTTGAACCAAGCA 17402  
QY 1419 GGTGAGGTTCAGTAAAGTGAATGCTGCTGTGCTTCCAGCTGGGCGACAGAGCA 1478  
DB 17401 GACAGAGGTTCAGTAAAGTGAATGCTGCTGTGCTTCCAGCTGGGCGACAGAGCG 17342  
QY 1479 AGACTGTGTCTCAGA 1493  
DB 17341 AAACGCCGTCTTAAAA 17327

RESULT 5  
US-10-060-332-3/C

Sequence 3, Application US/10060332

Patent No. 6528294

GENERAL INFORMATION:

APPLICANT: Li, Jiajin et al

TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC

TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES

FILE REFERENCE: THEROF

CURRENT APPLICATION NUMBER: US/10/060,332

CURRENT FILING DATE: 2002-02-01

NUMBER OF SEQ ID NOS: 4

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 3

LENGTH: 43950

TYPE: DNA

ORGANISM: Homo sapiens

US-10-060-332-3

Query Match 12.4%; Score 269.8; DB 4; Length 43950;

Best Local Similarity 77.9%; Pred. No. 3.1e-52;

Matches 339; Conservative 0; Mismatches 92; Indels 4; Gaps 1;

QY 1063 GCGAACCTGTGGACACGCTCATATGTTCTGAGTGTGGAGGTTAGAGAGGAGATGGC 1122  
DB 17761 GGGCACAAGTGGCACAACCTGTATCTCCGAACTTTGGAGGCTAGGTGGAGATGGC 17702  
QY 1123 TTGAACACAGGTGTTCAGGCGCAGCTGGCAACATTAACAGATCTGTCTCAAAAAA 1182  
DB 17701 TTGAGGCGAGAGTTTCAAGACCACTGGGCAACATAGACAGACCCCTATCTTAAAAA 17642  
QY 1183 AAAAAAAAAAAAAA---AAGAGAGAGGCGCGGCGTGTGTCTACGCTGTATCC 1238

DB 17641 AAAAAAAAAAAAAAAGAGAGAAATATAGTCAAGCGCGGTGTGTCTATGCTGTATCC 17582

QY 1239 CAGCACTTTGGAGGCGCGGCGGAGTCACTGTGTCAAGATTTGAGACCAAGCT 1298

DB 17581 CAGCACTTTGGAGGCGCGGCGGAGTCACTGTGTCAAGATTTGAGACCAAGCT 17522

QY 1299 GCGCAACATGGCAAAACCCCTGTGTACTCAAAATGCAAAATTTAGCCAGCGTGTAGC 1358

DB 17521 GACCAACATGGCAAAACCCCTGTGTACTCAAAATGCAAAATTTAGCTGGGCTGTAGC 17462

QY 1359 AGGCACTGTATCCAGCTTCTGTAGGAGCTGGAGGAGAGATCGTTGAACCAAGCA 1418

DB 17461 GGGTGTGTGTATCCAGCTTCTGTAGGAGCTGGAGGAGAGATCGTTGAACCAAGCA 17402

QY 1419 GGTGAGGTTCAGTAAAGTGAATGCTGCTGTGCTTCCAGCTGGGCGACAGAGCA 1478

DB 17401 GACAGAGGTTCAGTAAAGTGAATGCTGCTGTGCTTCCAGCTGGGCGACAGAGCG 17342

QY 1479 AGACTGTGTCTCAGA 1493

DB 17341 AAACGCCGTCTTAAAA 17327

RESULT 6  
US-09-798-096-10/C

Sequence 10, Application US/09798096

Patent No. 639378

GENERAL INFORMATION:

APPLICANT: Donna T. Ward

TITLE OF INVENTION: ANTISENSE MODULATION OF REG2 EXPRESSION

FILE REFERENCE: KTS-0207

CURRENT APPLICATION NUMBER: US/09/798,096

CURRENT FILING DATE: 2001-03-01

NUMBER OF SEQ ID NOS: 89

SEQ ID NO 10

LENGTH: 99500

TYPE: DNA

ORGANISM: Homo sapiens

FEATURES:

US-09-798-096-10

Query Match 12.4%; Score 269.2; DB 4; Length 99500;

Best Local Similarity 73.7%; Pred. No. 5.9e-52;

Matches 386; Conservative 0; Mismatches 103; Indels 35; Gaps 2;

QY 1039 GTGACAAAACAGATATTAAGTGGAAACCGTGGACACAGCTCATATGTTAGTGTCTT 1098

DB 43878 GGGAAAAAATAAAGTCTTGGTGAAGTGTGTGTCTATGACTGTATCCAGCCCTTT 43819

QY 1099 -GGAGGTTGAGAGGAGAGATGCTTGAACAAGTGTTCAGGCGCAGCTGGCAACA 1157

DB 43818 GGGAGGCGCAAGGAGAGATGCTTGAAGTTCAGAGAGTTCAGACAGCTGGGCAACA 43759

QY 1158 TAAACAAGTCTGTCTCTCAAAAAAATAAATAAATAAATAAATAAATAAATAAATAA 1206

DB 43758 CAGTGAAGCCTGTCTCTCAAAAAAATAAATAAATAAATAAATAAATAAATAAATAA 43699

QY 1207 -----GGGCGGCGGTGTGTGTCTCAAGCTGCTCAAGCTGTATATCCAGCA 1243

DB 43698 TTTTGT 43639

QY 1244 CTTTGGAGGCGGAGCGGCGGAGTCACTGTGTCAAGAGTTTGAACAGCTGGGCA 1303

DB 43638 CTTTGGAGGCGGAGAGAGTGTATCACTGTGTCAAGAGTTTGAACAGCTGGGCA 43579

QY 1304 ACATGCAAAACCCCTGTGTCTCAAAATGCAAAATTTAGCCAGCGTGTAGAGGCA 1363

DB 43578 ATGTGTGAAGAACCCGCTTTTCTAATAAATAAATAAATAAATAAATAAATAAATAA 43519

QY 1364 CTTGTATCCAGCTTCTTGGAGGCTGAGGAGAGATCGCTTGAACCAAGAGGTGG 1423



Db 43518 TGTGATCCAGACTCGGAGGCTGAGGACGAGAAATGCTTGAACCTGGAGGTGG 43459  
QY 1424 AGGTTCAGTAAGTGAAGATGTGCGCTTGCATCTCCAGCTGGGCGACAAAGCAAACT 1483  
Db 43458 AGGTTCAGTAAGTGAAGATGTGCGCTTGCATCTCCAGCTGGGCGACAAAGTGAAGCT 43399  
QY 1484 CTGTCTCAGAAAAAAGAGAGAGAGAGAGAGAGAG 1527  
Db 43398 TTGCTTCACAAAAAAGAGAGAGAGAGAGAGAGAG 43355

## RESULT 7

US-09-813-817-3/c  
; Sequence 3, Application US/09813817  
; Patent No. 6340583  
; GENERAL INFORMATION:  
; APPLICANT: YAN, Chunhua et al.  
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
; TITLE OF INVENTION: THEREOF  
; FILE REFERENCE: CL001178  
; CURRENT APPLICATION NUMBER: US/09/813,817  
; CURRENT FILING DATE: 2001-03-22  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 59065  
; TYPE: DNA  
; ORGANISM: Human  
US-09-813-817-3

Query Match 12.3%; Score 267.4; DB 4; Length 59065;  
Best Local Similarity 78.2%; Pred. No. 1.2e-51;  
Matches 351; Conservative 0; Mismatches 86; Indels 12; Gaps 2;

QY 1058 AAGCTGCGAACCCTGACACCTCATAGTTTACTGCTGGAGGTGAGAGAGAGG 1117  
Db 23369 AGGCCAGCATAGTGGCTCATCTATATCCACAGC-ATTGAGAGGCCAGAGAGAGG 23311  
QY 1118 ATGGCTTGAACACAGGTGTTCAAGGCGAGCTGGGCAACATACAAATCCTGCTCTCA 1177  
Db 23310 ACTGCTTGAAGCCAGAGTTTGAACACAGCTGGGTTAACAACAAAGCCTTATCTTAC 23251  
QY 1178 AAAAAAAAAAAAAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1237  
Db 23250 TAAAAAATCATGCAATCA-----GGCCAGTATGTTGCTCAGCGCTGTATTC 23202  
QY 1238 CCAGCATTTGGAGGCGAGCCGAGCGGAGTACCTGTGTCAGAGATTGAGACCAAGC 1297  
Db 23201 CCAGCATTTGGAGGCGTGAAGCGGAGCATCACTGAGGTTGAGAGTTGAGACTAGCC 23142  
QY 1298 TGGCCAACTGGCAAAACCCGCTGTGATCTCAAAATGCAAAATTAGCAGGCGGTAG 1357  
Db 23141 TGACCAACATGAGAAACCTGTCTTACTATAAAATPACAAAATTAGCAGGCGGTAG 23082  
QY 1358 CAGGCACTGTATCCAGCTACTTGGAGGCTGAGAGCGAGAAATCGTTGAACCCAGG 1417  
Db 23081 CAGTCTGTATCCAGCTACTTGGAGGCTGAGAGCGTGAACAGAAATCGTTGAACCCAGG 23022  
QY 1418 AGGTGAGGTGAGTAAGCTGAGATGTCGCTGTGACTCCAGCCTGGGCAACAAGGC 1477  
Db 23021 AGGCAAGGTGAGTGAAGCTGAGTTCGTCATTCAGCTGGGCAACAAGAGT 22962  
QY 1478 AAGACTGTCTTCAAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1506  
Db 22961 GAAACTGTCTCAAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 22933

RESULT 8  
US-09-978-197-3/c  
; Sequence 3, Application US/09978197  
; Patent No. 6403353  
; GENERAL INFORMATION:

; APPLICANT: YAN, Chunhua et al.  
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
; TITLE OF INVENTION: THEREOF  
; FILE REFERENCE: CL001178D1V  
; CURRENT APPLICATION NUMBER: US/09/978,197  
; CURRENT FILING DATE: 2001-10-17  
; PRIOR APPLICATION NUMBER: 09/813,817  
; PRIOR FILING DATE: 2001-03-22  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 59065  
; TYPE: DNA  
; ORGANISM: Human  
US-09-978-197-3

Query Match 12.3%; Score 267.4; DB 4; Length 59065;  
Best Local Similarity 78.2%; Pred. No. 1.2e-51;  
Matches 351; Conservative 0; Mismatches 86; Indels 12; Gaps 2;

QY 1058 AAGCTGCGAACCCTGACACCTCATAGTTTACTGCTGGAGGTGAGAGAGAGG 1117  
Db 23369 AGGCCAGCATAGTGGCTCATCTATATCCACAGC-ATTGAGAGGCCAGAGAGAGG 23311  
QY 1118 ATGGCTTGAACACAGGTGTTCAAGGCGAGCTGGGCAACATACAAATCCTGCTCTCA 1177  
Db 23310 ACTGCTTGAAGCCAGAGTTTGAACACAGCTGGGTTAACAACAAAGCCTTATCTTAC 23251  
QY 1178 AAAAAAAAAAAAAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1237  
Db 23250 TAAAAAATCATGCAATCA-----GGCCAGTATGTTGCTCAGCGCTGTATTC 23202  
QY 1238 CCAGCATTTGGAGGCGAGCCGAGCGGAGTACCTGTGTCAGAGATTGAGACCAAGC 1297  
Db 23201 CCAGCATTTGGAGGCGTGAAGCGGAGCATCACTGAGGTTGAGAGTTGAGACTAGCC 23142  
QY 1298 TGGCCAACTGGCAAAACCCGCTGTGATCTCAAAATGCAAAATTAGCAGGCGGTAG 1357  
Db 23141 TGACCAACATGAGAAACCTGTCTTACTATAAAATPACAAAATTAGCAGGCGGTAG 23082  
QY 1358 CAGGCACTGTATCCAGCTACTTGGAGGCTGAGAGCGAGAAATCGTTGAACCCAGG 1417  
Db 23081 CAGTCTGTATCCAGCTACTTGGAGGCTGAGAGCGTGAACAGAAATCGTTGAACCCAGG 23022  
QY 1418 AGGTGAGGTGAGTAAGCTGAGATGTCGCTGTGACTCCAGCCTGGGCAACAAGGC 1477  
Db 23021 AGGCAAGGTGAGTGAAGCTGAGTTCGTCATTCAGCTGGGCAACAAGAGT 22962  
QY 1478 AAGACTGTCTTCAAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1506  
Db 22961 GAAACTGTCTCAAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 22933

## RESULT 9

US-09-608-285A-8/c  
; Sequence 8, Application US/09608285A  
; Patent No. 6335013  
; GENERAL INFORMATION:  
; APPLICANT: Ford, John  
; APPLICANT: Mulero, Julio  
; APPLICANT: Yeung, George  
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO CD39-LIKE  
; TITLE OF INVENTION: POLYPEPTIDES  
; FILE REFERENCE: 28110/36570  
; CURRENT APPLICATION NUMBER: US/09/608,285A  
; CURRENT FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: 09/583,231  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 09/557,800  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/481,238  
; PRIOR FILING DATE: 2000-01-11











RESULT 13  
US-09-370-625A-8/c  
; Sequence 8, Application US/09370625A  
; Patent No. 6600052  
; GENERAL INFORMATION:  
; APPLICANT: Ford, John

	Query Match	12.2%	Score 265.6;	DB 4,	Length 9365;
	Best Local Similarity	79.0%	Pred. No. 1.6e-51;		
	Matches 297;	Conservative 34;	Mismatches 34;	Indels 11;	Gaps 2;
Qy	1143	CCAGCTGGGCACATACAGATCTGTCTCTCAAAAAAAAAAAAAAAAAAGAAAG-			1201
db	7802	CCAGATGGGTGCACGACCAAGACCCGTCTCAAAAAAAAAAAAAAAAAAG			7743



QY 1202 -----AGAGAGGCGCGGCTGTGATCAAGCTGTATCCAGACTTTGGAGG 1253  
DB 7742 AAMCAGCAAAAAAGGCTGGGTGTGTGCTCAATCTTATTCATCTTTGGAGG 7683  
QY 1254 CCGAGCGGCGCGGATCACTGTGTGAGAGTTTGAACAAGCTGGCAACATGGCAA 1313  
DB 7682 CCGAGGCGGTGTGATCACTGTGTGAGAGTTTGAAGAGAGCTGGCAACATGGCAA 7623  
QY 1314 ACCCGCTGTCTAATAATGC-AAAAATTAGCCAGCGGTGTGTGAGAGCACTGTAT 1371  
DB 7622 ACCCGCTGTCTAATAATGC-AAAAATTAGCCAGCGGTGTGTGAGAGCACTGTAT 7563  
QY 1372 CCGAGCTGTGTGAGAGCGGTGTGAGAGCACTGTATCCAGAGAGTGTGAGTGA 1431  
DB 7562 CCGAGCTGTGTGAGAGCGGTGTGAGAGCACTGTATCCAGAGAGTGTGAGTGA 7503  
QY 1432 GTAAGCTGAGATGTGCGCTGTGATCCAGCTGGGCGCAAGAGCAAGACTGTCTCA 1491  
DB 7502 GTGAGCGGTGATCTGTGAGAGTGTGATCCAGCTGGGCGCAAGAGCAAGACTGTCTCA 7443  
QY 1492 GAAAAAATATAATATAA 1507  
DB 7442 AAAAAAATATAATATAA 7427

## RESULT 14

US-09-608-285A-42/c  
Sequence 42, Application US/09608285A  
Patent No. 6335013  
GENERAL INFORMATION:  
APPLICANT: Ford, John  
APPLICANT: Mulero, Julio  
APPLICANT: Yeung, George  
TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO CD39-LIKE  
TITLE OF INVENTION: POLYPEPTIDES  
FILE REFERENCE: 28110/36570  
CURRENT APPLICATION NUMBER: US/09/608,285A  
CURRENT FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: 09/583,231  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 09/557,800  
PRIOR FILING DATE: 2000-04-25  
PRIOR APPLICATION NUMBER: 09/481,238  
PRIOR FILING DATE: 2000-01-11  
PRIOR APPLICATION NUMBER: 09/370,265  
PRIOR FILING DATE: 1999-08-09  
PRIOR APPLICATION NUMBER: PCT/US99/16180  
PRIOR FILING DATE: 1999-07-16  
PRIOR APPLICATION NUMBER: 09/350,836  
PRIOR FILING DATE: 1999-07-09  
PRIOR APPLICATION NUMBER: 09/273,447  
PRIOR FILING DATE: 1999-03-19  
PRIOR APPLICATION NUMBER: 09/244,444  
PRIOR FILING DATE: 1999-02-04  
PRIOR APPLICATION NUMBER: 09/122,449  
PRIOR FILING DATE: 1998-07-24  
PRIOR APPLICATION NUMBER: 09/118,205  
PRIOR FILING DATE: 1998-07-16  
NUMBER OF SEQ ID NOS: 60  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 42  
LENGTH: 14747  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (13641)  
OTHER INFORMATION: n = adenosine or guanine or cytosine or thymidine  
US-09-608-285A-42

Query Match 12.2%; Score 265.6; DB 4; Length 14747;  
Best Local Similarity 79.0%; Pred. No. 1.9e-51;  
Matches 297; Conservative 34; Mismatches 34; Indels 11; Gaps 2;

QY 1143 CCGAGCTGGCAACATTAACAAGATCTGTCTCAAAAAAAAAAAAAAAAAAGAAAG- 1201  
DB 11162 CCGAGCTGGGTGACAGAGCAACACCTGTCTCAAAAAAAAAAAAAAAAAAG 11103  
QY 1202 -----AGAGAGGCGCGGCTGTGATCAAGCTGTATCCAGACTTTGGAGG 1253  
DB 11102 AAMCAGCAAAAAAGGCTGGGTGTGTGCTCAATCTTATTCATCTTTGGAGG 11043  
QY 1254 CCGAGCGGCGCGGATCACTGTGTGAGAGTTTGAACAAGCTGGCAACATGGCAA 1313  
DB 11042 CCGAGGCGGTGTGATCACTGTGTGAGAGTTTGAAGAGAGCTGGCAACATGGCAA 10983  
QY 1314 ACCCGCTGTCTAATAATGC-AAAAATTAGCCAGCGGTGTGTGAGAGCACTGTAT 1371  
DB 10982 ACCCGCTGTCTAATAATGC-AAAAATTAGCCAGCGGTGTGTGAGAGCACTGTAT 10923  
QY 1372 CCGAGCTGTGTGAGAGCGGTGTGAGAGCACTGTATCCAGAGAGTGTGAGTGA 1431  
DB 10922 CCGAGCTGTGTGAGAGCGGTGTGAGAGCACTGTATCCAGAGAGTGTGAGTGA 10863  
QY 1432 GTAAGCTGAGATGTGCGCTGTGATCCAGCTGGGCGCAAGAGCAAGACTGTCTCA 1491  
DB 10862 GTGAGCGGTGATCTGTGAGAGTGTGATCCAGCTGGGCGCAAGAGCAAGACTGTCTCA 10803  
QY 1492 GAAAAAATATAATATAA 1507  
DB 10802 AAAAAAATATAATATAA 10787

## RESULT 15

US-09-557-800C-42/c  
Sequence 42, Application US/09557800C  
Patent No. 6476211  
GENERAL INFORMATION:  
APPLICANT: Ford, John  
APPLICANT: Mulero, Julio  
APPLICANT: Yeung, George  
TITLE OF INVENTION: Methods and Materials Relating to CD39-Like  
TITLE OF INVENTION: POLYPEPTIDES  
FILE REFERENCE: 28110/36457  
CURRENT APPLICATION NUMBER: US/09/557,800C  
CURRENT FILING DATE: 2000-04-25  
PRIOR APPLICATION NUMBER: 09/481,238  
PRIOR FILING DATE: 2000-01-11  
PRIOR APPLICATION NUMBER: 09/370,265  
PRIOR FILING DATE: 1999-08-09  
PRIOR APPLICATION NUMBER: PCT/US99/16180  
PRIOR FILING DATE: 1999-07-16  
PRIOR APPLICATION NUMBER: 09/350,836  
PRIOR FILING DATE: 1999-07-09  
PRIOR APPLICATION NUMBER: 09/273,447  
PRIOR FILING DATE: 1999-03-19  
PRIOR APPLICATION NUMBER: 09/244,444  
PRIOR FILING DATE: 1998-07-24  
PRIOR APPLICATION NUMBER: 09/122,449  
PRIOR FILING DATE: 1998-02-04  
PRIOR APPLICATION NUMBER: 09/118,205  
PRIOR FILING DATE: 1998-07-16  
NUMBER OF SEQ ID NOS: 56  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 42  
LENGTH: 14747  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (13641)  
OTHER INFORMATION: n = adenosine or guanine or cytosine or thymidine  
US-09-557-800C-42

Query Match 12.2%; Score 265.6; DB 4; Length 14747;  
Best Local Similarity 79.0%; Pred. No. 1.9e-51;  
Matches 297; Conservative 34; Mismatches 34; Indels 11; Gaps 2;







GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 12, 2004, 21:22:51 ; Search time 762.198 Seconds

(without alignments)  
10491.276 Million cell updates/sec

Title: US-09-966-880A-15

Perfect score: 2172  
Sequence: 1 cccctgataaggttgatga.....agatgatgaataactt 2172

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2432557 seqs, 184079884 residues

Total number of hits satisfying chosen parameters: 4865114

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA.\*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEM\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEM\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEM\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEM\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEM\_PUB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/2/pubpna/US10\_NEM\_PUB.seq:\*  
17: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*  
18: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	2172	100.0	2172	9	US-09-966-880A-15
2	2172	100.0	2818	9	US-09-966-880A-7
3	2172	100.0	6564	9	US-09-966-880A-10
4	2172	100.0	11204	9	US-09-966-880A-35
5	295.2	13.6	13862	10	US-09-764-891-5477
6	295.2	13.6	13862	10	US-09-764-891-10204
7	295.2	13.6	13862	10	US-10-205-428-1003
8	292.6	13.5	7809	10	US-09-764-891-6094
9	284.4	13.1	32249	10	US-09-764-891-7364
10	281.4	13.0	65608	9	US-09-962-436-292
11	281.4	13.0	65608	9	US-09-962-832-119
12	281.4	13.0	65608	9	US-09-954-531-180
13	280.2	12.9	11821	9	US-09-764-877-2857
14	280.2	12.9	11821	15	US-10-242-515-2857
15	280	12.9	15297	13	US-10-003-295-3

16	278.2	12.8	32204	10	US-09-764-891-8537	Sequence 8537, App
17	277.4	12.8	30000	14	US-10-262-552-33	Sequence 33, App1
18	276.6	12.7	37940	15	US-10-348-072A-12	Sequence 12, App1
19	276.6	12.7	38000	15	US-10-175-627-11	Sequence 11, App1
20	276.4	12.7	176080	12	US-10-235-192A-43	Sequence 43, App1
21	275.6	12.7	160361	12	US-10-235-192A-35	Sequence 35, App1
22	275.4	12.7	3966	9	US-09-764-855-227	Sequence 227, App
23	275.4	12.7	3966	14	US-10-072-345-227	Sequence 227, App
24	275	12.7	6482	14	US-10-180-410-6	Sequence 6, App1
25	275	12.7	20050	14	US-10-180-410-5	Sequence 5, App1
26	274.8	12.7	15362	9	US-09-764-877-2856	Sequence 2856, App
27	274.8	12.7	15362	15	US-10-242-515-2856	Sequence 2856, App
28	274.2	12.6	11811	10	US-09-728-552-7	Sequence 7, App1
29	274.2	12.6	176001	16	US-10-210-556-27	Sequence 27, App1
30	274.2	12.6	186739	16	US-10-210-556-19	Sequence 43, App1
31	274	12.6	4963	9	US-09-764-877-2903	Sequence 2903, App
32	274	12.6	4963	15	US-10-242-515-2903	Sequence 2903, App
33	273.6	12.6	174424	9	US-09-967-768A-314	Sequence 314, App
34	273.6	12.6	174424	10	US-09-960-706-969	Sequence 969, App
35	273	12.6	2401	15	US-10-027-632-102635	Sequence 102635, App
36	273	12.6	26657	9	US-09-810-673A-3	Sequence 3, App1
37	273	12.6	26657	14	US-10-395-242-3	Sequence 3, App1
38	272	12.5	17491	14	US-10-017-161-1995	Sequence 1995, App
39	270.8	12.5	670	15	US-10-027-632-201375	Sequence 201375, App
40	270.8	12.5	1061	14	US-10-277-161-66	Sequence 66, App1
41	270.8	12.5	107820	10	US-09-792-616-1	Sequence 1, App1
42	270.2	12.4	17286	9	US-09-764-877-3234	Sequence 3234, App
43	270.2	12.4	17286	15	US-10-242-515-3234	Sequence 3234, App
44	269.8	12.4	43950	13	US-10-060-323-3	Sequence 3, App1
45	269.8	12.4	43950	14	US-10-339-657-3	Sequence 3, App1

#### ALIGNMENTS

RESULT 1  
US-09-966-880A-15  
Sequence 15, Application US/09966880A  
Patent No. US20020164743A1  
GENERAL INFORMATION:  
APPLICANT: Honjo, Yasuko  
APPLICANT: Muramatsu, Masamichi  
TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
FILE REFERENCE: 06501-088001  
CURRENT APPLICATION NUMBER: US/09/966, 880A  
PRIOR FILING DATE: 2001-09-28  
PRIOR APPLICATION NUMBER: PCT/JP00/01918  
PRIOR FILING DATE: 2000-03-28  
PRIOR APPLICATION NUMBER: JP 11-371382  
PRIOR FILING DATE: 1999-12-27  
PRIOR APPLICATION NUMBER: JP 11-178999  
PRIOR FILING DATE: 1999-06-24  
PRIOR APPLICATION NUMBER: JP 11-87192  
PRIOR FILING DATE: 1999-03-29  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 15  
LENGTH: 2172  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-966-880A-15  
Query Match 100.0%; Score 2172; DB 9; Length 2172;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 CCCCTGATAGGTTGATGACTTACGAGAGCATTTGCGATCTTGGACCTTATAGCA 60  
DB 1 CCCCTGATAGGTTGATGACTTACGAGAGCATTTGCGATCTTGGACCTTATAGCA 60  
OY 61 CTTCCAGAGATGTCACACAGATGATATCTTGTGTAAGACAGTGAATAAACAAGT 120  
DB 61 CTTCCAGAGATGTCACACAGATGATATCTTGTGTAAGACAGTGAATAAACAAGT 120



QY 121 CTTCAAGTCTTCTGTGTTTATTTCTTCAACTCTCACTTTCTAGAGTTACAGAAA 180  
 DB 121 CCTCAAGTCTTCTGTGTTTATTTCTTCAACTCTCACTTTCTAGAGTTACAGAAA 180  
 QY 181 AATTATATACGACTCTTTAAAGATCTATGCTTGAAGAAATAGAGAAAGAAACAGCT 240  
 DB 181 AATTATATACGACTCTTTAAAGATCTATGCTTGAAGAAATAGAGAAAGAAACAGCT 240  
 QY 241 CTGGCAGGAGCGTGTCAATTTGATGAGTTTGAATGCAACTTGTCCCTTCTGGA 300  
 DB 241 CTGGCAGGAGCGTGTCAATTTGATGAGTTTGAATGCAACTTGTCCCTTCTGGA 300  
 QY 301 ATTAAGAACTGAGAGAGCTGGAGAGATCTTAAAGTGTCAAGTTTCTATGACTTTTA 360  
 DB 301 ATTAAGAACTGAGAGAGCTGGAGAGATCTTAAAGTGTCAAGTTTCTATGACTTTTA 360  
 QY 361 GGTAGAGATGAGAGAGAGATCTTAAAGAGATGAGAGAGATCAAAATGTTTTA 420  
 DB 361 GGTAGAGATGAGAGAGAGATCTTAAAGAGATGAGAGAGATCAAAATGTTTTA 420  
 QY 421 TATCAACATCTTATATTTGATTCATTTGATTAAGTGTGTTAGTATGATTTT 480  
 DB 421 TATCAACATCTTATATTTGATTCATTTGATTAAGTGTGTTAGTATGATTTT 480  
 QY 481 TCTATTCCTTCCCTTGAAGTTTACTTCAAGTAAACAAACTCTTCATGAGCCATGA 540  
 DB 481 TCTATTCCTTCCCTTGAAGTTTACTTCAAGTAAACAAACTCTTCATGAGCCATGA 540  
 QY 541 TCTATTCCTTCCCTTGAAGTTTACTTCAAGTAAACAAACTCTTCATGAGCCATGA 600  
 DB 541 TCTATTCCTTCCCTTGAAGTTTACTTCAAGTAAACAAACTCTTCATGAGCCATGA 600  
 QY 601 GATTAATATCCAAATGAGCTGTATGTTTAAATCAGAGAGAGATGTTTATGTTG 660  
 DB 601 GATTAATATCCAAATGAGCTGTATGTTTAAATCAGAGAGAGATGTTTATGTTG 660  
 QY 661 TACAAAAGAGATTTGATGAGTGGAGATGAGATGAGATGAGATGAGATGAGATGAG 720  
 DB 661 TACAAAAGAGATTTGATGAGTGGAGATGAGATGAGATGAGATGAGATGAGATGAG 720  
 QY 721 TACAAAAGAGATTTGATGAGTGGAGATGAGATGAGATGAGATGAGATGAGATGAG 780  
 DB 721 TACAAAAGAGATTTGATGAGTGGAGATGAGATGAGATGAGATGAGATGAGATGAG 780  
 QY 781 TGGGTTGATGCTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 840  
 DB 781 TGGGTTGATGCTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 840  
 QY 841 AATTGAGAAACCCCAAACTTCAATATCAATATGAGAAACCAATGAGAGAGTGG 900  
 DB 841 AATTGAGAAACCCCAAACTTCAATATCAATATGAGAAACCAATGAGAGAGTGG 900  
 QY 901 CTTGAATGTTGGGAGAGAGAAATCTATTGGCTCTGTGGGCTCTTCAATCAGAAATG 960  
 DB 901 CTTGAATGTTGGGAGAGAGAAATCTATTGGCTCTGTGGGCTCTTCAATCAGAAATG 960  
 QY 961 CCAATCAGGTCAGGTTTGTCTATTTTGTATGTTGTGTGTGTGTGTGTGTGTGTGTGT 1020  
 DB 961 CCAATCAGGTCAGGTTTGTCTATTTTGTATGTTGTGTGTGTGTGTGTGTGTGTGTGT 1020  
 QY 1021 TAACTATATAGAGATGAGCAAAAGATGATTAAGTGAAGTGGAGAGTGGAGAGTGG 1080  
 DB 1021 TAACTATATAGAGATGAGCAAAAGATGATTAAGTGAAGTGGAGAGTGGAGAGTGG 1080  
 QY 1081 TCAATAGTTAGCTGTGGAGGTTGAGAGAGAGAGATGAGTGAACAGAGTGTGA 1140  
 DB 1081 TCAATAGTTAGCTGTGGAGGTTGAGAGAGAGAGATGAGTGAACAGAGTGTGA 1140  
 QY 1141 GGCAGAGCTGGGCAACATTAAGATCTCTCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1200  
 DB 1141 GGCAGAGCTGGGCAACATTAAGATCTCTCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1200  
 QY 1200 GGCAGAGCTGGGCAACATTAAGATCTCTCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1260  
 DB 1200 GGCAGAGCTGGGCAACATTAAGATCTCTCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1260

QY 1201 GAGAGAGGCGGGGCTGTGAGCTCAAGCTGTATCCAGACCTTTGGAGGCGGAGCC 1260  
 DB 1201 GAGAGAGGCGGGGCTGTGAGCTCAAGCTGTATCCAGACCTTTGGAGGCGGAGCC 1260  
 QY 1261 GGGCGGATCACTGTGTGTGAGAGGTTTGAAGCAGAGCTGTGAGAGAGAGAGAGAGAG 1320  
 DB 1261 GGGCGGATCACTGTGTGTGAGAGGTTTGAAGCAGAGCTGTGAGAGAGAGAGAGAGAG 1320  
 QY 1321 CTGTACTGAATATGAATAAATAGCAGGCGGTGTGAGAGAGAGAGAGAGAGAGAGAGAG 1380  
 DB 1321 CTGTACTGAATATGAATAAATAGCAGGCGGTGTGAGAGAGAGAGAGAGAGAGAGAGAG 1380  
 QY 1381 TTGGAGAGCTGAG 1440  
 DB 1381 TTGGAGAGCTGAG 1440  
 QY 1441 GATGTCGCGGTGCACTCCAGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500  
 DB 1441 GATGTCGCGGTGCACTCCAGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500  
 QY 1501 AAAAAAG 1560  
 DB 1501 AAAAAAG 1560  
 QY 1561 TGCAGAGAAATGTGCTTATATCAACAAATATGAAGAGAGAGAGAGAGAGAGAGAGAGAG 1620  
 DB 1561 TGCAGAGAAATGTGCTTATATCAACAAATATGAAGAGAGAGAGAGAGAGAGAGAGAGAG 1620  
 QY 1621 TCTCTTTGAGTGTCTATTTTGTCTTAACTGTCTTGAAGAGAGAGAGAGAGAGAGAGAG 1680  
 DB 1621 TCTCTTTGAGTGTCTATTTTGTCTTAACTGTCTTGAAGAGAGAGAGAGAGAGAGAGAG 1680  
 QY 1681 ATACCATATCCCTGTGCGGTATTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740  
 DB 1681 ATACCATATCCCTGTGCGGTATTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740  
 QY 1741 ACAGAGAACTGAAAG 1800  
 DB 1741 ACAGAGAACTGAAAG 1800  
 QY 1801 AGAGTTAAATATGTTACTTCAATGATGATTAATTTATTAATTTATTTATTTATTTATTT 1860  
 DB 1801 AGAGTTAAATATGTTACTTCAATGATGATTAATTTATTAATTTATTTATTTATTTATTT 1860  
 QY 1861 AATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1920  
 DB 1861 AATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1920  
 QY 1921 AATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1980  
 DB 1921 AATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1980  
 QY 1981 GGTGCTACGAGAGCAATTTCTGATTTTATGATTTTATGATTTTATGATTTTATGATTTTAT 2040  
 DB 1981 GGTGCTACGAGAGCAATTTCTGATTTTATGATTTTATGATTTTATGATTTTATGATTTTAT 2040  
 QY 2041 TGGCTCACTTCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 2100  
 DB 2041 TGGCTCACTTCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 2100  
 QY 2101 TACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2160  
 DB 2101 TACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2160  
 QY 2161 GGAATTAATTTG 2172  
 DB 2161 GGAATTAATTTG 2172



GENERAL INFORMATION:  
APPLICANT: Honjo, Tasuku  
TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
FILE REFERENCE: 06501-088001  
CURRENT APPLICATION NUMBER: US/09/966,880A  
CURRENT FILING DATE: 2001-09-28  
PRIOR APPLICATION NUMBER: PCT/JP00/01918  
PRIOR FILING DATE: 2000-03-28  
PRIOR APPLICATION NUMBER: JP 11-373382  
PRIOR FILING DATE: 1999-12-27  
PRIOR APPLICATION NUMBER: JP 11-178999  
PRIOR FILING DATE: 1999-06-24  
PRIOR APPLICATION NUMBER: JP 11-87192  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 7  
LENGTH: 2818  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (80)...(673)  
FEATURE:  
NAME/KEY: 5'UTR  
LOCATION: (1)...(79)  
FEATURE:  
NAME/KEY: 3'UTR  
LOCATION: (677)...(2818)  
US-09-966-880A-7

Query Match 100.0%; Score 2172; DB 9; Length 2818;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCCTGATGAGTGTGATGACTTACGAGACGATTCCTTCTTGGACCTTGTATAGCA 60  
DB CCCCTGATGAGTGTGATGACTTACGAGACGATTCCTTCTTGGACCTTGTATAGCA 682  
QY 61 CTTCCAGGATGTCAACACGATGAAATATCTCTGCTGAAGACAGTGTATAAAAACGT 120  
DB CTTCCAGGATGTCAACACGATGAAATATCTCTGCTGAAGACAGTGTATAAAAACGT 742  
QY 121 CCTTAAGTCTCTGTTTTTATCTCACTCTCACTTCTTGAAGTTACAGAAAA 180  
DB CCTTAAGTCTCTGTTTTTATCTCACTCTCACTTCTTGAAGTTACAGAAAA 802  
QY 181 AATATTAATACGACTCTTTAAAAAGATCTATGCTTGAATAAGAGAGAGACACAGT 240  
DB AATATTAATACGACTCTTTAAAAAGATCTATGCTTGAATAAGAGAGAGACACAGT 862  
QY 241 CTGGCCAGGAGCGTCTGCAATGTGTGACGTTTGAATGCAACATGTGCCCTACTGGA 300  
DB CTGGCCAGGAGCGTCTGCAATGTGTGACGTTTGAATGCAACATGTGCCCTACTGGA 922  
QY 261 GGTAGGATGAGAGAGAGAGTATCTTAAAAAGATGATGAGAGATCAATGTTTTTA 420  
DB GGTAGGATGAGAGAGAGAGTATCTTAAAAAGATGATGAGAGATCAATGTTTTTA 1042  
QY 301 ATTAACAGAACTGACAGACCTGGAGACATCTAAAGTGCACGTTTTTCTATGACTTTA 360  
DB ATTAACAGAACTGACAGACCTGGAGACATCTTAAAGTGCACGTTTTTCTATGACTTTA 982  
QY 361 GGTAGGATGAGAGAGAGTATCTTAAAAAGATGATGAGAGATCAATGTTTTTA 420  
DB GGTAGGATGAGAGAGAGTATCTTAAAAAGATGATGAGAGATCAATGTTTTTA 1042  
QY 421 TATCAACATCTTTATTTATTTGATTCATTTGATTAGCAAGTGTGTATAGATTTT 480  
DB TATCAACATCTTTATTTATTTGATTCATTTGATTAGCAAGTGTGTATAGATTTT 1102  
QY 481 TGTATTTCTTTTCTTGGACGTTTACTTTCAAGTAAACAAATCTTCCATCAGGCGATGA 540  
DB TGTATTTCTTTTCTTGGACGTTTACTTTCAAGTAAACAAATCTTCCATCAGGCGATGA 1162  
QY 541 TCTATAGACCTCTTAATGAGATCTGGGTGATTTGACCCCAACCATCTCTCCAA 600

DB 1163 TCTATAGACCTCTTAATGAGATCTGGGTGATTTGACCCCAACCATCTCTCCAA 1222  
QY 601 GCATTATATCCATATATGCGTGTATGTTTAAATCAGAGAAAGATTTTATGTTTG 660  
DB GCATTATATCCATATATGCGTGTATGTTTAAATCAGAGAAAGATTTTATGTTTG 1282  
QY 661 TCAAAAAGAAATGTTATGGGTGGGAGATGAGATATAGACATCATGTGACCTTCA 720  
DB TCAAAAAGAAATGTTATGGGTGGGAGATGAGATATAGACATCATGTGACCTTCA 1342  
QY 721 GCTACTTTATTAAGATCTTAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780  
DB GCTACTTTATTAAGATCTTAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1402  
QY 781 TGGGTGATGTCTGAAGTGAAGTCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 840  
DB TGGGTGATGTCTGAAGTGAAGTCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1462  
QY 841 AATTTGAAACACCCCAAACTTCAATATCAATTAATGCAAAATGGAAGAGAGTTG 900  
DB AATTTGAAACACCCCAAACTTCAATATCAATTAATGCAAAATGGAAGAGAGTTG 1522  
QY 901 CTTGAATGTGGGAGAGAGAAATCTATGCTCTGCTGCTCTCTCAATCTCAGAAATG 960  
DB CTTGAATGTGGGAGAGAGAAATCTATGCTCTGCTGCTCTCTCAATCTCAGAAATG 1582  
QY 961 CCAATCAGTCAAGGTTTCTCAATTTTGTATGTGTGTATGCTTCCCAAGGTATAT 1020  
DB CCAATCAGTCAAGGTTTCTCAATTTTGTATGTGTGTATGCTTCCCAAGGTATAT 1642  
QY 1021 TAACTATATAGAGAGTGTGACAAAACAGATATTAAGTGCAGACCTGTGACACGC 1080  
DB TAACTATATAGAGAGTGTGACAAAACAGATATTAAGTGCAGACCTGTGACACGC 1702  
QY 1081 TCAATTTCTAGTCTGTTGGAGGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140  
DB TCAATTTCTAGTCTGTTGGAGGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1762  
QY 1141 GGCACGCTGGGCAATTAACAAGATCTGTCTCTCAAAAAAAGAGAGAGAGAGAGAG 1200  
DB GGCACGCTGGGCAATTAACAAGATCTGTCTCTCAAAAAAAGAGAGAGAGAGAGAG 1822  
QY 1201 GAGAGAGGCGGCGCTGTGTGCTCAAGCTGTATCCAGACATTTGGAGGCGGAGGC 1260  
DB GAGAGAGGCGGCGCTGTGTGCTCAAGCTGTATCCAGACATTTGGAGGCGGAGGC 1882  
QY 1261 GGGCGGATCACTGTGTGAGAGATTTGAGACCAAGCTGTGCAAAATGCAAAACCCCGT 1320  
DB GGGCGGATCACTGTGTGAGAGATTTGAGACCAAGCTGTGCAAAATGCAAAACCCCGT 1942  
QY 1321 CTTGACTCAAAATGCAAAATTAAGCAGGCGGTGTAGAGAGAGAGAGAGAGAGAGAG 1380  
DB CTTGACTCAAAATGCAAAATTAAGCAGGCGGTGTAGAGAGAGAGAGAGAGAGAGAG 2002  
QY 1381 TTGGAGGCTGAGAGAGAGATTCGCTGAACCCAGAGAGTGAAGTTGACATTAAGCTGA 1440  
DB TTGGAGGCTGAGAGAGAGATTCGCTGAACCCAGAGAGTGAAGTTGACATTAAGCTGA 2062  
QY 1441 GATCGTGCCTTCACTCCAGCTGGGCAACAAGAGAGAGAGAGAGAGAGAGAGAGAG 1500  
DB GATCGTGCCTTCACTCCAGCTGGGCAACAAGAGAGAGAGAGAGAGAGAGAGAGAG 2122  
QY 1501 AAAAAAG 1560  
DB AAAAAAG 2182  
QY 1561 TGCAGAGAAATGTGCTTATATCAACAATATGTAAGAGAGAGAGAGAGAGAGAGAGAG 1620  
DB TGCAGAGAAATGTGCTTATATCAACAATATGTAAGAGAGAGAGAGAGAGAGAGAGAG 2242  
QY 1621 TCTCTTTGCTGTATTTGTCTTAACAATGTCTTTGACAGTGAAGAAAAATTTCA 1680







1141 GGCACGCTGGGCAACATACAGATCCCTCTCTCAAAAAAAAAAAAAAAAAAGAA 1200  
Db GGCACGCTGGGCAACATACAGATCCCTCTCTCAAAAAAAAAAAAAAAAAAGAA 4940  
QY 1201 GAGAGAGGCGCGCGCTGGCTCAAGCTGTAAATCCAGACATTTGGAGGCGGAGCC 1260  
Db GAGAGAGGCGCGCGCTGGCTCAAGCTGTAAATCCAGACATTTGGAGGCGGAGCC 5000  
QY 1261 GGGCGGATACCTGTGTGTGAGAGTTGAGACCAAGCTGGCCAAATGGCAAAACCCGCT 1320  
Db GGGCGGATACCTGTGTGTGAGAGTTGAGACCAAGCTGGCCAAATGGCAAAACCCGCT 5060  
QY 1321 CTGTACTCAAAATGCAAAATTTAGACGAGCGTGGTAGAGGACCTGTAAATCCAGTAC 1380  
Db CTGTACTCAAAATGCAAAATTTAGACGAGCGTGGTAGAGGACCTGTAAATCCAGTAC 5120  
QY 1381 TTGGAGGCTGAGGAGAGAAATCGCTTGAACCCAGAGGTGAGAGTTGCAAGTAACTGA 1440  
Db TTGGAGGCTGAGGAGAGAAATCGCTTGAACCCAGAGGTGAGAGTTGCAAGTAACTGA 5180  
QY 1441 GATCGTCCGTGCACTCCAGCCTGGGCGACAGAGCAAGACTGTGTCTCAGAAAAAAA 1500  
Db GATCGTCCGTGCACTCCAGCCTGGGCGACAGAGCAAGACTGTGTCTCAGAAAAAAA 5240  
QY 1501 AAAAAAAGAT 1560  
Db AAAAAAAGAT 5300  
QY 1561 TGCAAGAAATGTGTTTATCCAAACAAATGTAAGAGCCATTAAGGATCCCTAATTTG 1620  
Db TGCAAGAAATGTGTTTATCCAAACAAATGTAAGAGCCATTAAGGATCCCTAATTTG 5360  
QY 1621 TCTCTTTGATGTCTATTTGTCCCTAACAACTGCTTTGACAGTGAAGAAAAATATTCGA 1680  
Db TCTCTTTGATGTCTATTTGTCCCTAACAACTGCTTTGACAGTGAAGAAAAATATTCGA 5420  
QY 1681 ATACCAATATCCCTGTGCGGTATTAAGTACCAAGCCCTTGCAATGAGTGAAGATCC 1740  
Db ATACCAATATCCCTGTGCGGTATTAAGTACCAAGCCCTTGCAATGAGTGAAGATCC 5480  
QY 1741 ACGGAAATCTGAATGCAACACTGTCTAATTTTAATCTAATGTAAGTAAATTTGTA 1800  
Db ACGGAAATCTGAATGCAACACTGTCTAATTTTAATCTAATGTAAGTAAATTTGTA 5540  
QY 1801 AGAGTAAATATGTACTTCAATGATTAATTTATTTATTTATTTATTTATTTATTTATG 1860  
Db AGAGTAAATATGTACTTCAATGATTAATTTATTTATTTATTTATTTATTTATTTATG 5600  
QY 1861 ATTTTATTTAATGATATTTCTTTCTGATATTAATTAATGAGTCTCAAGCTTCATA 1920  
Db ATTTTATTTAATGATATTTCTTTCTGATATTAATTAATGAGTCTCAAGCTTCATA 5660  
QY 1921 AATTATTAATTTAGAAATGATTTCTAATCAACAGTATTAATTTGTAATCTGAGTAT 1980  
Db AATTATTAATTTAGAAATGATTTCTAATCAACAGTATTAATTTGTAATCTGAGTAT 5720  
QY 1981 GGTGCTACGAGGCAATTTCTGTGATTTTATGTAATCTTTATGACAGCAATTTGCTTC 2040  
Db GGTGCTACGAGGCAATTTCTGTGATTTTATGTAATCTTTATGACAGCAATTTGCTTC 5780  
QY 2041 TGGCTCACTTTCATGATTAATAATATGATTAATTTGAAAGCTGTGAAGTAA 2100  
Db TGGCTCACTTTCATGATTAATAATATGATTAATTTGAAAGCTGTGAAGTAA 5840  
QY 2101 TACCAATATAATATAATATAAGTATTAATTAATGAAGTAAATATAATATAGATAT 2160  
Db TACCAATATAATATAATATAAGTATTAATTAATGAAGTAAATATAATATAGATAT 5900  
QY 2161 GGAATTAATCTG 2172  
Db GGAATTAATCTG 5912

RESULT 4  
US-09-966-880a-35  
; Sequence 35, Application US/09966880A  
; Patent No. US20020164743A1  
; GENERAL INFORMATION:  
; APPLICANT: Honjo, Taaku  
; APPLICANT: Muramatsu, Masamichi  
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
; FILE REFERENCE: 06501-088001  
; CURRENT APPLICATION NUMBER: US/09/966, 880A  
; PRIOR FILING DATE: 2001-09-28  
; PRIOR APPLICATION NUMBER: PCT/JP00/01918  
; PRIOR FILING DATE: 2000-03-28  
; PRIOR APPLICATION NUMBER: JP 11-371382  
; PRIOR FILING DATE: 1999-12-27  
; PRIOR APPLICATION NUMBER: JP 11-178999  
; PRIOR FILING DATE: 1999-06-24  
; PRIOR APPLICATION NUMBER: JP 11-87192  
; PRIOR FILING DATE: 1999-03-29  
; NUMBER OF SEQ. ID NOS: 36  
; SOFTWARE: Fastsq for Windows Version 4.0  
; SEQ. ID NO: 35  
; LENGTH: 11204  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-966-880a-35  
Query Match 100.0%; Score 2172; DB 9; Length 11204;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CCCTGTATAGAGTTGATGACTTACGAGAGCAATTTGATCTTGGACTTGTATGCA 60  
Db 8956 CCCTGTATAGAGTTGATGACTTACGAGAGCAATTTGATCTTGGACTTGTATGCA 9015  
QY 61 CTTCAGAGATGTCAACAGATGAATATCTGCTGAAGACAGTGAATAAAAACAGT 120  
Db 9016 CTTCAGAGATGTCAACAGATGAATATCTGCTGAAGACAGTGAATAAAAACAGT 9075  
QY 121 CCTCAAGCTCTGCTGTTTATTTATCTTCAACTCTCACTTTCTTAAGTTTACAGAAAA 180  
Db 9076 CCTCAAGCTCTGCTGTTTATTTATCTTCAACTCTCACTTTCTTAAGTTTACAGAAAA 9135  
QY 181 ATATTATATACGACTCTTTTAAAAAGATCTATGCTTTGAAATATGAGAGAAACACAGT 240  
Db 9136 ATATTATATACGACTCTTTTAAAAAGATCTATGCTTTGAAATATGAGAGAAACACAGT 9195  
QY 241 CTGGCCAGGAGAGTGTGCAATTTGTGCAATTTTGAATGCAATGTCCCTACCTGGGA 300  
Db 9196 CTGGCCAGGAGAGTGTGCAATTTGTGCAATTTTGAATGCAATGTCCCTACCTGGGA 9255  
QY 301 ATAAACAGACTGACGAGCCTGGAGCATCTTAAGTGTCAAGCTTTTCTATGACTTTTA 360  
Db 9256 ATAAACAGACTGACGAGCCTGGAGCATCTTAAGTGTCAAGCTTTTCTATGACTTTTA 9315  
QY 361 GGTAGATGAGAGAGAGTATCTTAAAAAGCATGTGTGAGAGATCAATGTTTAA 420  
Db 9316 GGTAGATGAGAGAGAGTATCTTAAAAAGCATGTGTGAGAGATCAATGTTTAA 9375  
QY 421 TATCAACATCTTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 480  
Db 9376 TATCAACATCTTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 9435  
QY 481 TCTATTTCTTCCCTGAGGTTTACTTCAAGTAAACAAACCTTCAATCAGGCGCATGA 540  
Db 9436 TCTATTTCTTCCCTGAGGTTTACTTCAAGTAAACAAACCTTCAATCAGGCGCATGA 9495  
QY 541 TCTATAGACCTCTCAATGAGATATCTGGGTATTTGAGCCCAACCAATCTCTCCAA 600  
Db 9496 TCTATAGACCTCTCAATGAGATATCTGGGTATTTGAGCCCAACCAATCTCTCCAA 9555  
QY 601 GCATTATATTCATCATGCGCTGTATGTTTATATGAGAGAAAGATTTTATGTTTG 660



Db	9556	GCATTATATCCAAATCATGCGCTGTATGTTTTTAATCAGAGAAGAAGATGTTTTTAATGTTTG	95.15
Qy	661	TACAAAAGAAAGATTGTTATGSGTGGGGAATGAGAGTATAGACATGCATGTCACCTTCAA	720
Db	9616	TACAAAAGAAAGATTGTTATGSGTGGGGAATGAGAGTATAGACATGCATGTCACCTTCAA	96.75
Qy	721	GCATCTTATTAAGGATCTTAAATGGGACAGAGACGTGAACAGACACCTTAATTA	780
Db	9676	GCTACTTTAATTAAGATCTTAAATGGGACAGAGACGTGAACAGACACCTTAATTA	9735
Qy	781	TGGGTGATGTCTGAAGTAGCAAAATTTCTGAAAAGCAAACTCTTTTAAGAAATGTCCT	840
Db	9736	TGGGTGATGTCTGAAGTAGCAAAATTTCTGAAAAGCAAACTCTTTTAAGAAATGTCCT	9795
Qy	841	AATTTAGAAACCCCAAACTTCAATATCATATATTAGCAAACAATTGAAAGAAATTG	900
Db	9796	AATTTAGAAACCCCAAACTTCAATATCATATATTAGCAAACAATTGAAAGAAATTG	9855
Qy	901	CTTGAATGTGGGGAAGGAAATTCATATGGCTCTGTGGGTCTCTCATCTCAGAAATG	960
Db	9856	CTTGAATGTGGGGAAGGAAATTCATATGGCTCTGTGGGTCTCTCATCTCAGAAATG	9915
Qy	961	CCAATCAGGTCAAGGTTGCTACATTTTGTATGTGTATGCTCTCCCAAAAGTATAT	1020
Db	9916	CCAATCAGGTCAAGGTTGCTACATTTTGTATGTGTATGCTCTCCCAAAAGTATAT	9975
Qy	1021	TAACTATATAAGAGAGTTGTGACAAACAGATGATTAAGCTGCGAACCTGTGCACAGC	1080
Db	9976	TAACTATATAAGAGAGTTGTGACAAACAGATGATTAAGCTGCGAACCTGTGCACAGC	10035
Qy	1081	TCATATGTTTACGTCTTTGGGAGGTTGAGAGGAGAGATGGCTTGAACACAGGTTCAA	1140
Db	10036	TCATATGTTTACGTCTTTGGGAGGTTGAGAGGAGAGATGGCTTGAACACAGGTTCAA	10095
Qy	1141	GGCCAGCCTGGGCAACATAACAAGATCCTGTCTCTCAAAAAAAAAAAAAAAAAAGAA	1200
Db	10096	GGCCAGCCTGGGCAACATAACAAGATCCTGTCTCTCTCAAAAAAAAAAAAAAAAAAGAA	10155
Qy	1201	GAGAGAGGCGCGGCGTGTGGCTCAGCGCTGTATTCACAGCATTTGGGAGGCGGAGCC	1260
Db	10156	GAGAGAGGCGCGGCGTGTGGCTCAGCGCTGTATTCACAGCATTTGGGAGGCGGAGCC	10215
Qy	1261	GGGCGGATCACCTGTGTCAAGAGTTTGAACCAAGCTGTGCCAACATGGCAAAACCCGT	1320
Db	10216	GGGCGGATCACCTGTGTCAAGAGTTTGAACCAAGCTGTGCCAACATGGCAAAACCCGT	10275
Qy	1321	CTGTATCAAAATGCAAAATTTAGCCAGGCGGTGTAGCAGGACCTGTATCCACAGCTAC	1380
Db	10276	CTGTATCAAAATGCAAAATTTAGCCAGGCGGTGTAGCAGGACCTGTATTCACAGCTAC	10335
Qy	1381	TTGGAGGCTGAGCAGGAGAAATGCTTGAACCCAGAGGTGAGGTTGACATTAAGCTGA	1440
Db	10336	TTGGAGGCTGAGGAGGAGAAATGCTTGAACCCAGAGGTGAGGTTGACATTAAGCTGA	10395
Qy	1441	GATCGTCCGTTTGCACTTCAGCTGTGGCGCAAAAGCAACACTGTCTTCAGAAAAAAA	1500
Db	10396	GATCGTCCGTTTGCACTTCAGCTGTGGCGCAAAAGCAACACTGTCTTCAGAAAAAAA	10455
Qy	1501	AAAAAAGAT	1560
Db	10456	AAAAAAGAT	10515
Qy	1561	TGCAAGAGAAATTGCTTATTCAACAAAATGTAAAGAGCCAAATAAGATCCCTATTG	1620
Db	10516	TGCAAGAGAAATTGCTTATTCAACAAAATGTAAAGAGCCAAATAAGATCCCTATTG	10575
Qy	1621	TCTCTTTGGTGTCTAATTTGTCCCTAACACTGTCTTTGACAGTGAAGAAAAAATTCAGA	1680
Db	10576	TCTCTTTGGTGTCTAATTTGTCCCTAACACTGTCTTTGACAGTGAAGAAAAAATTCAGA	10635
Qy	1681	AATAACCATATCCCTGTGCGTTATTAATCAACACCTGTGCATGAAGATGAGACAGATCC	1740
Db	10636	AATAACCATATCCCTGTGCGTTATTAATCAACACCTGTGCATGAAGATGAGACAGATCC	10695

QY	1741	ACAGGAAACTGGAATGCAACGTCTTATTTTAACTGATATGACATTAAGTTGTAA	1800
Db	10696	ACAGGAAACTGGAATGCAACGTCTTATTTTAACTGATATGACATTAAGTTGTAA	10755
QY	1801	AGAGTTAAAAATGTTTACTTCAATGATTCATTTATATTTATATTTTGGCTAATG	1860
Db	10756	AGAGTTAAAAATGTTTACTTCAATGATTCATTTATATTTATATTTTGGCTAATG	10815
QY	1861	ATTTTATTAATGATTTTCCTTTCTGATATATTTGAATGGAGTCCAAAGCTTCA	1920
Db	10816	ATTTTATTAATGATTTTCCTTTCTGATATATTTGAATGGAGTCCAAAGCTTCA	10875
QY	1921	AATTTATACCTTAGAATATGATTCATTAACAAGATGATAATGTAACTTGACGAT	1980
Db	10876	AATTTATACCTTAGAATATGATTCATTAACAAGATGATAATGTAACTTGACGAT	10935
QY	1981	GGTGTCACGAAGCCATTTCTCTTGATTTTATTAACCTTTATGACAGAAATTTGCTC	2040
Db	10936	GGTGTCACGAAGCCATTTCTCTTGATTTTATTAACCTTTATGACAGAAATTTGCTC	10995
QY	2041	TGGTCACTTCATCAGTTAATAATGATAAATATTTTGAAGCTGTGAGATAA	2100
Db	10996	TGGTCACTTCATCAGTTAATAATGATAAATATTTTGAAGCTGTGAGATAA	11055
QY	2101	TACCAATTAATAATATTAATAAGTATTTATATGAACTTAAATTAATAAATACGATGAT	2160
Db	11056	TACCAATTAATAATAATTAATAAGTATTTATATGAACTTAAATTAATAAATACGATGAT	11115
QY	2161	GGAATAAAGCTTG 2172	
Db	11116	GGAATAAAGCTTG 11127	

```

RESULT 5
US-09-764-891-5477
/ Sequence 5477, Application US/09764891
/ Publication No. US20030077808A1
/ GENERAL INFORMATION:
/ APPLICANT: Rosen et al.
/ TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
/ FILE REFERENCE: PC006
/ CURRENT FILING DATE: 2001-01-17
/ Prior application data removed - consult PALM or file wrapper
/ NUMBER OF SEQ ID NOS: 10231
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 5477
/ LENGTH: 13862
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-09-764-891-5477

Query Match      13.6%; Score 295.2; DB 10; Length 13862;
Best Local Similarity 74.6%; Pred. No. 3.5e-51;
Matches 397; Conservative 0; Mismatches 133; Indels 2; Gaps 2;

Dh      1045 AAACAGAAATGATAAAGCTGCCAACCGTGGCACACGCTCATATGTTCTAGCTCCTTGGAGG 1104
1508 AAACATTAATAAGCTGGCTGGCTGCTCATGCTGCTATGCCATGCCACACTTTGGGGG 1567

Qy      1105 TTGAGGAGGAGGATAGCGCTTGTAACACAGGTGTTCAAGGCCAGCCCTGGGCAATACAAAG 1164
1568 CTGAGGAGGAGGAGTCACTTGAAGTCAGAGATTGAGACAGCCTTGGCCAAATGGCGAA 1622

Dh      1165 ATCTGTCTCTCAAAAATAAAAATAAAAATAAAAATGAAAGA-GAGAGGCGCGGCGTGTGGC 1223
1628 ACCCATCTCACTTAAAAATCAAAAGTTAAGCAGGCAATGGGCGGGGATGTTGGGTGGC 1687

Qy      1224 TCAGCGCTGTATCCACAGCACTTTGGAGGCCGAGCCGGGCGAATCACTGTGTCAAGA 1283
1688 TCATCGCTGTATCCACAGCACTTTGGAGGCCGAGGTGGGCGGATCACTGAGGTCAAGA 1747

```



QY 1284 GTTGAAGACAGCTGGCCACATGGCAAAACCCGCTGTACTCAAAATGCAAAATTA 1343  
DB 1748 GTTCAAGACCAAGCTGTGACCAACATGGCAAAACCCGCTGTACTCAAAATTA 1807  
QY 1344 GCCAGCGGTGTGACAGGCACTGTATCCAGCTACTTGGAGGCTGAGGAGAGAAAT 1403  
DB 1808 GCCAGCGTTATGTGTGGGCGCTTATATCCAGCTACTGGAAGGCTGAGGAGAGAAAT 1867  
QY 1404 CGCTTAACCCAGAGAGAGTGGAGTGGCAGTAAGCTGAGATGCGCGTGGCACTCCAGCC 1463  
DB 1868 CGTTTAAGCCAGAGAGCGAGAGTTGGACCCAGAGCTGTGCGCACTGCACTCCAGCC 1927  
QY 1464 TGGGCGACCAAGCAAGCAAGCTGTCTCAGAAAAAAGAGAGAGAGAGAGAA 1523  
DB 1928 TGGGCGCATAA-AGCGAGACTCTGTCTCAGAAAAAAGAGAGAGAGAGAA 1986  
QY 1524 AGAGAAACAATATTGGGAGAGAAAGATGGGGAAGCAATTGCAAGAAATTGTG 1575  
DB 1987 AGAAAAAATTAGCCAGCGTGTGTGGCATGCACTGTGTCTAGCTACTTGGG 2038

## RESULT 6

US-09-764-891-10204  
; Sequence 10204, Application US/09764891  
; Publication No. US20030077808A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC006  
; CURRENT APPLICATION NUMBER: US/09/764,891  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION DATA REMOVED - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 10231  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 10204  
; LENGTH: 13862  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-891-10204

Query Match 13.6%; Score 295.2; DB 10; Length 13862;  
Best Local Similarity 74.6%; Pred. No. 3.5e-51;  
Matches 397; Conservative 0; Mismatches 133; Indels 2; Gaps 2;

QY 1045 AAACAGATGATAAAGCTGCAACCGTGGCAACGCTCATAGTCTAGCTGTGGAGG 1104  
DB 1508 AAACCTTAATAGCTGTGGCTGTGTGCTCATGCTGTATCCAGACACTTTGGAGG 1567  
QY 1105 TTGAGAGAGAGAGATGGCTTGAACACAGGTGTTCAAGGCCAGCTGGGCAATACAG 1164  
DB 1568 CTGAGGACAGAGAGATCACTTGAGTCAAGAGTTTGAAGCAACCTGGCCAACTAGCG 1627  
QY 1165 ATCTGTCTCTCAAAAAAAGAAAAAAGAAAGA-GAGAGGCGCGGCGTGTGGC 1223  
DB 1628 ACCCATCTCTCTATAAAATACAAAGTTAGCCAGGCAATGGCGCGGATGTGGTGGC 1687  
QY 1224 TCAGCGCTGTATCCAGCACTTTGGAGGCGGAGCGGCGGATCACTGTGTGAGG 1283  
DB 1688 TCATGCTGTATCCAGCACTTTGGAGGCGGAGCGGCGGATCACTGTGTGAGG 1747  
QY 1284 GTTGAAGACAGCTGGCCACATGGCAAAACCCGCTGTACTCAAAATGCAAAATTA 1343  
DB 1748 GTTCAAGACCAAGCTGTGACCAACATGGCAAAACCCGCTGTACTCAAAATTA 1807  
QY 1344 GCCAGCGGTGTGACAGGCACTGTATCCAGCTACTTGGAGGCTGAGGAGAGAAAT 1403  
DB 1808 GCCAGCGTTATGTGTGGGCGCTTATATCCAGCTACTGGAAGGCTGAGGAGAGAAAT 1867  
QY 1404 CGCTTAACCCAGAGAGTGGAGTGGCAGTAAGCTGAGATGCGCGTGGCACTCCAGCC 1463  
DB 1868 CGTTTAAGCCAGAGAGCGAGAGTTGGACCCAGAGCTGTGCGCACTGCACTCCAGCC 1927  
QY 1464 TGGGCGACCAAGCAAGCAAGCTGTCTCAGAAAAAAGAGAGAGAGAGAA 1523

DB 1928 TGGGCGATTA-AGCGAGACTCTGTCTCAGAAAAAAGAGAGAGAGAGAA 1986  
QY 1524 AGAGAAACAATATTGGGAGAGAAAGATGGGGAAGCAATTGCAAGAAATTGTG 1575  
DB 1987 AGAAAAAATTAGCCAGCGTGTGTGGCATGCACTGTGTCTAGCTACTTGGG 2038

## RESULT 7

US-10-205-428-1003  
; Sequence 1003, Application US/10205428  
; Publication No. US20030108907A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PA117C1  
; CURRENT APPLICATION NUMBER: US/10/205,428  
; PRIOR FILING DATE: 2002-07-26  
; PRIOR APPLICATION NUMBER: 09/764,892  
; PRIOR FILING DATE: 2001-01-17  
; PRIOR APPLICATION NUMBER: 60/179,065  
; PRIOR FILING DATE: 2000-01-31  
; PRIOR APPLICATION NUMBER: 60/180,628  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: 60/214,886  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: 60/217,487  
; PRIOR FILING DATE: 2000-07-11  
; PRIOR APPLICATION NUMBER: 60/225,758  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/220,963  
; PRIOR FILING DATE: 2000-07-26  
; PRIOR APPLICATION NUMBER: 60/217,496  
; PRIOR FILING DATE: 2000-07-11  
; PRIOR APPLICATION NUMBER: 60/225,447  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/218,290  
; PRIOR FILING DATE: 2000-07-14  
; REMAINING PRIOR APPLICATION DATA REMOVED - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 1019  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1003  
; LENGTH: 13862  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-205-428-1003

Query Match 13.6%; Score 295.2; DB 14; Length 13862;  
Best Local Similarity 74.6%; Pred. No. 3.5e-51;  
Matches 397; Conservative 0; Mismatches 133; Indels 2; Gaps 2;

QY 1045 AAACAGATGATAAAGCTGCAACCGTGGCAACGCTCATAGTCTAGCTGTGGAGG 1104  
DB 1508 AAACCTTAATAGCTGTGGCTGTGTGCTCATGCTGTATCCAGACACTTTGGAGG 1567  
QY 1105 TTGAGAGAGAGAGATGGCTTGAACACAGGTGTTCAAGGCCAGCTGGGCAATACAG 1164  
DB 1568 CTGAGGACAGAGATCACTTGAGTCAAGAGTTTGAACCAAGCTGGCCAACTAGCG 1627  
QY 1165 ATCTGTCTCTCAAAAAAAGAAAAAAGAAAGA-GAGAGGCGCGGCGTGTGGC 1223  
DB 1628 ACCCATCTCTCTATAAAATACAAAGTTAGCCAGGCAATGGCGCGGATGTGGTGGC 1687  
QY 1224 TCAGCGCTGTATCCAGCACTTTGGAGGCGGAGCGGCGGATCACTGTGTGAGG 1283  
DB 1688 TCATGCTGTATCCAGCACTTTGGAGGCGGAGCGGCGGATCACTGTGTGAGG 1747  
QY 1284 GTTGAAGACAGCTGGCCACATGGCAAAACCCGCTGTACTCAAAATGCAAAATTA 1343  
DB 1748 GTTCAAGACCAAGCTGTGACCAACATGGCAAAACCCGCTGTACTCAAAATTA 1807  
QY 1344 GCCAGCGGTGTGACAGGCACTGTATCCAGCTACTTGGAGGCTGAGGAGAGAAAT 1403



Db 1808 GCCAGCCTTATGTTAGGCGCCTATATATCCAGACTCTGGAGGCTGAGCAGAGAAAT 1867  
Qy 1404 CGCTTGAACCCAGAGAGTGGAGGTTGCACTGAAGCTGAGATGTGCGCTTGCATCCAGCC 1463  
Db 1868 CGTTTGAAGCCAGAGGCGGAGTTTGCAACCGAGCCGAGATTTGTGCACTGCATCCAGCC 1927  
Qy 1464 TGGGCGACAGAGCAAGACTGTCTCAGAAAAAAGAAAAAAGAGAGAGAGAGAA 1523  
Db 1928 TGGGCGAATAA-AGCAGAGACTGTCTCAGAAAAAAGAAAAAAGAAAAAAGAAAAA 1986  
Qy 1524 AGAGAACATATTTTGGAGAGAGAGATGGGAGGATTCAGAGAAATTTGTG 1575  
Db 1987 AGAAAAAATTAGCCAGCGGTGTGGCATGACCTGTAGTCTTAGTACTTGGG 2038

## RESULT 8

US-09-764-891-6094  
; Sequence 6094, Application US/09764891  
; Publication No. US20030077808A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC006  
; CURRENT APPLICATION NUMBER: US/09/764,891  
; PRIOR APPLICATION DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 10231  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 6094  
; LENGTH: 7809  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-891-6094

Query Match 13.5%; Score 292.6; DB 10; Length 7809;  
Best Local Similarity 77.2%; Pred. No. 9.3e-51;  
Matches 385; Conservative 0; Mismatches 104; Indels 10; Gaps 2;

Qy 1044 AAAAAAGATGATTAAGTGGCGAACCCTGAGCAGCCTCATATGTTCTAGCTCTTGGAG 1103  
Db 1081 AAAAAAGTAAATCAAGCTGGGCACTGTGTGCTCAGCCCTATATATCCAGCACTTTGGAG 1140  
Qy 1104 GTTGAAGAGGAGATGGCTTGAACAAGGTGTTCAAGGCCAGCTGGGCAACATPACAA 1163  
Db 1141 GCTGAAGGAGGAGATCACTTGAAGCCAGAGCTCAAGACCAACCTGCACAAGATGATGA 1200  
Qy 1164 GATCTGTCTCTCAAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG 1223  
Db 1201 AACCTGTCTCTCAATTTAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAA 1251  
Qy 1224 TCACGCTGTATATCCAGCACTTTGGAGGCGAGCGGAGGATCACTGTGTGAGCA 1283  
Db 1252 TCACCTGTATATCCAGCACTTTGGAGGCGAGCGGAGGATCACTGTGTGAGCA 1311  
Qy 1284 GTTGAAGAGGAGCTGGCACAATGGCAAAACCCGCTGTGTAATCAAAAAATGCAAAATTA 1343  
Db 1312 GTTCAAGAGAGGCTGACCAACATGTGAAACCCCAATGTCTACAAAAATTAATAAATA 1371  
Qy 1344 GCCAGGCTGTGAGCAGGCACTGTAAATCCAGCTACTTGGAGGCTGAGGCGAGAGAT 1403  
Db 1372 GCTGGGCTGTGTGCTCATGCTGTAAATCCAGCTACTTGGAGGCTGAGGCGAGAGAT 1431  
Qy 1404 CGCTTGAACCCAGAGAGTGGAGGTTGCACTGAAGCTGAGATGTGCGCTTGCATCCAGCC 1463  
Db 1432 CGATTGAACCTAGAGGCAAGGTTGTAGTGAGTGAATGTGCACTTGCATCCAGAC 1491  
Qy 1464 TGGGCGACAGAGG-AGAAGCTGTCTCAGAAAAAAGAAAAAAGAGAGAGAGAGAA 1522  
Db 1492 TGGGCGACAGAGCAAAAGCACTCACTCTCAAAAAAAGAAAAAAGAAAAAAGAAAAA 1551  
Qy 1523 AAGAGAACATATTTGGGA 1541  
Db 1552 GTTAAACAAGTATGGCA 1570

## RESULT 9

US-09-764-891-7364  
; Sequence 7364, Application US/09764891  
; Publication No. US20030077808A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC006  
; CURRENT APPLICATION NUMBER: US/09/764,891  
; PRIOR APPLICATION DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 10231  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 7364  
; LENGTH: 32249  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-891-7364

Query Match 13.1%; Score 284.4; DB 10; Length 32249;  
Best Local Similarity 75.7%; Pred. No. 9.9e-49;  
Matches 365; Conservative 0; Mismatches 116; Indels 1; Gaps 1;

Qy 1060 GCTGCAACCGTGGCAACGCTCATATGTTCTAGCTCTTGGAGGTTGAGAGAGAGAT 1119  
Db 15040 GCTGTGACAGTGTGCTCATGCTCTGTAACTCAAGCACTTGGAGATCTAGGCGGTGAGT 15099  
Qy 1120 GCTTGAACAGAGTGTCAAGGCCAGCTGGCAACATPACAAATCTGTCTCAAA 1179  
Db 15100 CGCTTGAAGCCAGAGATTCAGACCAAGCTGGCAACATAGTAGAGCCATCTPAAA 15159  
Qy 1180 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAA 1239  
Db 15160 AATTAATAATATTAATAAAGCTTTAGGCGCAGGCGGTGTGCTCAGCCTGTATCC 15219  
Qy 1240 AGCATTGGAGAGCGGCGGCGGAGATCACTGTGTGAGAGTTGAGACCAAGCTG 1299  
Db 15220 AGCCCTTGGAGAGCGCAAGGAGGCGGTGATTTGAGTGAAGTTTGAACCAAGCTG 15279  
Qy 1300 GCCAAGATGGCAAAACCCGCTGTACTCAAAATGCAAAATTAAGCCAGGCGGTGAGCA 1359  
Db 15280 GCCAAGATGGTGAATCTCATCTCTACTGAAATTAACCAAGCATGTGAGGCG 15339  
Qy 1360 GGCACCTGTATATCCAGCTACTTGGAGGCTGAGGAGAAATCGCTTGAATCCAGAG 1419  
Db 15340 GCGCCTGTATATCCAGCTACTCAAGAGATGAGGAGAGAAATGCTGAATCCAGAG 15399  
Qy 1420 GTGAGGTTGAGTGAAGTGAATGTCGCTGTGACTCCAGGCTGGGCGCAAGAGCA 1479  
Db 15400 GAGGACATGTGAGTGAAGCCAGAGTGTGCGCACTGCAAGCTGGGTAC-AGAGCCA 15458  
Qy 1480 GACTGTGTCTCAGAAAAAAGAAAAAAGAAAAAAGAGAGAGAGAGAAATATTTGG 1539  
Db 15459 GAGTCAATCTCAAAAAAAGAAAAAAGAAAAAAGAAATTAATTAATTAATTAATTA 15518  
Qy 1540 GA 1541  
Db 15519 TA 15520

## RESULT 10

US-09-962-436-292  
; Sequence 292, Application US/09962436  
; Patent No. US20020081301A1  
; GENERAL INFORMATION:  
; APPLICANT: Soppet, Daniel  
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Sigma  
; FILE REFERENCE: 689290-75  
; CURRENT APPLICATION NUMBER: US/09/962,436  
; CURRENT FILING DATE: 2001-09-25



PRIOR APPLICATION NUMBER: US/60/235,082  
PRIOR FILING DATE: 2000-09-25  
PRIOR APPLICATION NUMBER: US/60/234,924  
PRIOR FILING DATE: 2000-09-25  
NUMBER OF SEQ ID NOS: 568  
SOFTWARE: Patent in version 3.0  
SEQ ID NO 292  
LENGTH: 65608  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: n=a,t,g or c  
US-09-962-436-292

Query Match 13.0%; Score 281.4; DB 9; Length 65608;  
Best Local Similarity 81.2%; Pred. No. 6e-48;  
Matches 376; Conservative 0; Mismatches 81; Indels 6; Gaps 4;

QY 1060 GCTGCAACCGTGGCAACGCTCATAGTTCTAGCTGCTTGGAGGTTGAGAGGAGAT 1119  
DB 43089 GCCGGGTGTGTGGCTCAAGCTGTATCCAGCACTTTGGAGTCTGAGGAGAGAGAT 43148  
QY 1120 GGCTTGAACAGAGTTTCAAGGCGCTGGGCAACATTAACAAGATCTGTCTCTCAA 1179  
DB 43149 CCTTGAAGCCAGAAATTCAA-GCCAGCTGGGCAATGTAGGAGAGCGCATCTCTACTA 43207  
QY 1180 AAAAAAAAAAAAAAAAAAAGA--AAGAGAGAGGCGCGGGGTGTGCTCAAGCTGTATC 1237  
DB 43208 AAAAAAAAAAAAAAAAAAAGAGAGAAATTAGCCCGGGGTGTGCTCAAGCTGTATC 43267  
QY 1238 CCAGCACTTTGGAGGCGCGAGCCGGGCGGATCACTGTGTGAGAGATTGAGACCAACC 1297  
DB 43268 CCAGCACTTTGGAGGCGTGAAGGAGGCGGATCAC--GAGGTCAAGAGTTTGAACACAGCC 43325  
QY 1298 TGGCCAACTGGCAAAACCCGCTGTGTACTCAAAATGCAAAATTAAGCCAGCGGTGTAG 1357  
DB 43326 TGGCCAAAGTGTGAAACCCCTGTCTTACTTAAATAATTAAGCCAGCGGTGTAG 43385  
QY 1358 GAGGACCTGTATCCAGTACTTGGAGGCTGAGGAGAGAGATCGTTGAACCCAGG 1417  
DB 43386 CGGCGCTGTATCCAGTACTTGAAGGCTGAGGAGAGAGATTTGTTCAACCTGGG 43445  
QY 1418 AGGTGAGGTTGCAATAGCTGAGATCGTCCGTTGCACTCCAGCTGGGCGACAAAGAC 1477  
DB 43446 AGCGGAGGTTGCAAGTGAAGCCAGATCGTGCCACTGTACTCAGGCTGGGTGACAA-AGC 43504  
QY 1478 AAGACTCTGTCTCAGAAAAAAAAAAAAAAAAAGAGAGAGAGA 1520  
DB 43505 AAGACTCCATCTCAAAAAAAAAAAAAAAAAAGAAAGAAA 43547

RESULT 11  
US-09-962-832-119  
Sequence 119, Application US/09962832  
Patent No. US2002010821A1  
GENERAL INFORMATION:  
APPLICANT: Emnet, Reinhard  
TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu  
FILE REFERENCE: 689290-74  
CURRENT APPLICATION NUMBER: US/09/962,832  
PRIOR FILING DATE: 2001-09-25  
PRIOR APPLICATION NUMBER: US/60/235,077  
PRIOR FILING DATE: 2000-09-25  
PRIOR APPLICATION NUMBER: US/60/235,280  
PRIOR FILING DATE: 2000-09-25  
NUMBER OF SEQ ID NOS: 259  
SOFTWARE: Patent in version 3.0  
SEQ ID NO 119  
LENGTH: 65608  
TYPE: DNA  
ORGANISM: Homo sapiens

FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: n=a,t,g or c  
US-09-962-832-119

Query Match 13.0%; Score 281.4; DB 9; Length 65608;  
Best Local Similarity 81.2%; Pred. No. 6e-48;  
Matches 376; Conservative 0; Mismatches 81; Indels 6; Gaps 4;

QY 1060 GCTGCAACCGTGGCAACGCTCATAGTTCTAGCTGCTTGGAGGTTGAGAGGAGAT 1119  
DB 43089 GCCGGGTGTGTGGCTCAAGCTGTATCCAGCACTTTGGAGTCTGAGGAGAGAGAT 43148  
QY 1120 GGCTTGAACAGAGTTTCAAGGCGCTGGGCAACATTAACAAGATCTGTCTCTCAA 1179  
DB 43149 CCTTGAAGCCAGAAATTCAA-GCCAGCTGGGCAATGTAGGAGAGCGCATCTCTACTA 43207  
QY 1180 AAAAAAAAAAAAAAAAAAAGA--AAGAGAGAGGCGCGGGGTGTGCTCAAGCTGTATC 1237  
DB 43208 AAAAAAAAAAAAAAAAAAAGAGAGAAATTAGCCCGGGGTGTGCTCAAGCTGTATC 43267  
QY 1238 CCAGCACTTTGGAGGCGCGAGCCGGGCGGATCACTGTGTGAGAGATTGAGACCAACC 1297  
DB 43268 CCAGCACTTTGGAGGCGTGAAGGAGGCGGATCAC--GAGGTCAAGAGTTTGAACCAAGCC 43325  
QY 1298 TGGCCAACTGGCAAAACCCGCTGTGTACTCAAAATGCAAAATTAAGCCAGCGGTGTAG 1357  
DB 43326 TGGCCAAAGTGTGAAACCCCTGTCTTACTTAAATAATTAAGCCAGCGGTGTAG 43385  
QY 1358 GAGGACCTGTATCCAGTACTTGGAGGCTGAGGAGAGAGATCGTTGAACCCAGG 1417  
DB 43386 CGGCGCTGTATCCAGTACTTGAAGGCTGAGGAGAGAGATTTGTTCAACCTGGG 43445  
QY 1418 AGGTGAGGTTGCAATAGCTGAGATCGTCCGTTGCACTCCAGCTGGGCGACAAAGAC 1477  
DB 43446 AGCGGAGGTTGCAAGTGAAGCCAGATCGTGCCACTGTACTCAGGCTGGGTGACAA-AGC 43504  
QY 1478 AAGACTCTGTCTCAGAAAAAAAAAAAAAAAAAGAGAGAGAGA 1520  
DB 43505 AAGACTCCATCTCAAAAAAAAAAAAAAAAAAGAAAGAAA 43547

RESULT 12  
US-09-964-531-180  
Sequence 180, Application US/09954531  
Patent No. US20020165180A1  
GENERAL INFORMATION:  
APPLICANT: Weaver, Zoe  
TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Ca  
FILE REFERENCE: 689290-77  
CURRENT APPLICATION NUMBER: US/09/954,531  
PRIOR FILING DATE: 2002-05-02  
PRIOR APPLICATION NUMBER: US/60/233,133  
PRIOR FILING DATE: 2000-09-18  
PRIOR APPLICATION NUMBER: US/60/234,009  
PRIOR FILING DATE: 2000-09-20  
PRIOR APPLICATION NUMBER: US/60/234,034  
PRIOR FILING DATE: 2000-09-20  
PRIOR APPLICATION NUMBER: US/60/234,509  
PRIOR FILING DATE: 2000-09-22  
PRIOR APPLICATION NUMBER: US/60/234,567  
PRIOR FILING DATE: 2000-09-22  
NUMBER OF SEQ ID NOS: 1392  
SOFTWARE: Patent in version 3.0  
SEQ ID NO 180  
LENGTH: 65608  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)...(65608)  
OTHER INFORMATION: n=a,t,g or c



US-09-954-531-180

Query Match	13.0%;	Score 281.4;	DB 9;	Length 65608;
Best Local Similarity	81.2%;	Pred. No. 6e-46;		
Matches 376;	Conservative	0;	Mismatches 81;	Indels 6;
				Gaps 4;

QY	1060	GCCTGCMAACGGTGGCACACGCTCA	TAAGTTAGCTGTTGGAGGTTGAGAGAGAGAT	1119
Db	43089	GCCGGGTGTGTGGCTACGCTCTGTAAT	CCAGCATTTTGAAAGCTGTAGGACGAGGAT	43148
QY	1120	GGCTTGAACA	CGAGTGTTCAGGCGACCTGGGCAATACAAATCCTGTCTCAAA	1179
Db	43149	CCCTTGAAGCCAGAAAGTTCAA	-GCCACCTGGGCAATGTGAGAGACGCACATCTTCACTA	43207
QY	1180	AAAAAAAAAAAAAAAAA	GA--MAGAGAGAGGCGGGCGTGTGGCTACAGCCTGTAAATC	1237
Db	43208	AAAAAAAAAAAAAAAAA	AGAGAAATTTAGCCGGGCGTGTGGCTACAGCCTGTAAATC	43267
QY	1238	CCAGCACTTTGGAGAGCGCCGAGCCGGGCGGATCA	CTGTGTGTCAAGAGTTTGAGACCAAGCC	1297
Db	43268	CCAGCACTTTGGAGAGCGCTGAGGCAAGGGAATCA	-GAGTTCAGAGAGTTTGAGACCAAGCC	43325
QY	1298	TGGCCAAATGCGCAAAACCCCGCTTGTATCTCAAAATTC	CAAAATTTAGCCAGGCGTGTAG	1357
Db	43326	TGGCCAAATGTGAAACCCCGTGTCTATCAAAATTC	CAAAATTTAGCCAGGCGATGATG	43385
QY	1358	CAGGCACCTGTAAATCCGACGTA	CTTGGAGGCTGAGCGAAGAAATCGTTGAACCAAG	1417
Db	43386	CGGGCGCTGTAAATCCAGGTA	CTCAGAGGCTAGGCAAGAAATTCCTTCAACCTGGG	43445
QY	1418	AGGTGAGGTTGCAAGTAAGCTGAAGATGTGCGCTTG	CATTCACAGCTCTGGGCGCAAGAGAC	1477
Db	43446	AGGCGAAGTTGCAAGTGAAGCCAAATGCTGTCA	CTGTCCACAGCCTGGGTGACAA-AGC	43504
QY	1478	AAAGCTGTCTCAGAAAAA	AAAAAAAAAAGAGAGAGAGA	1520
Db	43505	AAAGCTCATCTCAAAAAA	AAAAAAAAAAGAGAGAGAA	43547

```

RESULT 13
US-09-764-877-2857
; Sequence 2857, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; PRIOR FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: Patencin Ver. 2.0
; SEQ. ID NO 2857
; LENGTH: 11821
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-877-2857

```

Query Match	12.9%	Score 280.2	DB 9	Length 11821
Best Local Similarity	74.4%	Pred. No. 4.5e-48		
Matches 386	Conservative	0	Mismatches 118	Indels 15
			Gaps	2
Qy	1013	AGGTATATTAACATATATTAAGACAGTGTGTACAAAACAGATGATTAAGCTGGGAACCTGT	10722	
Db	4599	AGGAACCTGAACTATCTGTGGCAGCAGCAAGAAATGAACATGAAGCAGCATGATG	46588	
Qy	1073	GCAACAGCTCAATGTTCTAGCTGCTTGGAGGTTGAGAGGAGAGATGCTTGAAACAAG	11323	
Db	4659	GCTCAGACCGCGTATCCCAAGCACTTAGA-ATTTAGGAGAGAGATCACTTGAAGCCAG	4717	
Qy	1133	GTTTCAAGGCGCAGCCGGGCAATTAACAAGATCTGTCTCAAAAAAAAAAAAAAAAAA	11921	
Db	4718	GAGTTCAAGACCACTCTGGGCAATATATGTATGACCTTGTCTCAAAAAAAAAAGAAAAA	4776	

QY	1193	AAAAAGAAAGAGAGAGGAGCCGGGAGCTGGTGGCTACGCGCTGTATCCAGACTTTGGGAG	1252
	4777	-----TTGGCAGAGCGTGTGGCTACCGCTGTATTCAGCACTTTGGGAG	4823
Db			
QY	1253	GCCGAGACGGGCGGATCACTGTGTGAGAGTTTGAGACAGCGTCGGCCAAATGGCAA	1312
	4824	GCCGAAAGGGGCAAGTTATCTGAGGTGAGAGTTGAGACAGCGTCGGCCAAATGGCAA	4883
Db			
QY	1313	AACCCCGTCTGTACTCAAAATGCAAAATTAGCCAGGCGTGTAGCAGGCACTGTATC	1372
	4884	AAOCCGGGTCTACTAAAAATATCAAAATTAGCTGGGCGTGGGCGCACCTGTATTC	4943
Db			
QY	1373	CCAGCTACTTGGAGAGGCTGAGGCGAGGAATTCGCTTGAACCCAGAGGTGAGGTTGAG	1433
	4944	CCAGCTACTTGGAGAGGCTGAGGCGAGGAATCACTTGAATTGGGAGGTGAGGTTGAG	5003
Db			
QY	1433	TAAAGTGAATCTGTGCGGTGCACTCCAGGCTGGGCGACAAGAGCAAGACTGTCTGAG	1492
Db	5004	TGAGCGGGATTCGACCACTTACACTGACGCTGGGCGAAGAGGTGAACCTGTCTCAA	5066
QY	1493	AAAAAAAAAAAAAGAGAGAGAGAGAAAGGAA	1531
Db	5064	AAATAAATATAATATAATATAATATAATATAATATA 5102	

```

RESULT 14
US-10-242-515-2857
Sequence 2857, Application US/10242515
Publication No. US20040009488A1
GENERAL INFORMATION:
APPLICANT: Rosen et al..
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC005C1
CURRENT APPLICATION NUMBER: US/10/242,515
CURRENT FILING DATE: 2002-09-13
PRIORITY APPLICATION NUMBER: 09/764,877
PRIORITY FILING DATE: 2001-01-17
PRIORITY APPLICATION NUMBER: 60/119,065
PRIORITY FILING DATE: 2000-01-31
PRIORITY APPLICATION NUMBER: 60/180,628
PRIORITY FILING DATE: 2000-02-04
PRIORITY APPLICATION NUMBER: 60/214,886
PRIORITY FILING DATE: 2000-06-28
PRIORITY APPLICATION NUMBER: 60/217,487
PRIORITY FILING DATE: 2000-07-11
PRIORITY APPLICATION NUMBER: 60/225,758
PRIORITY FILING DATE: 2000-08-14
PRIORITY APPLICATION NUMBER: 60/220,963
PRIORITY FILING DATE: 2000-07-26
PRIORITY APPLICATION NUMBER: 60/217,496
PRIORITY FILING DATE: 2000-07-11
PRIORITY APPLICATION NUMBER: 60/225,447
PRIORITY FILING DATE: 2000-08-14
PRIORITY APPLICATION NUMBER: 60/218,290
PRIORITY FILING DATE: 2000-07-14
Remaining Prior Application data removed - See File Wrapper or PALM
NUMBER OF SEQ ID NOS: 4031
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2857
LENGTH: 11821
TYPE: DNA
ORGANISM: Homo sapiens
US-10-242-515-2857

```

	Query Match	Score	280.2;	DB	Length	11821;
	Best Similarity	74.48;				
	Matches	386;	Conservative	0;	Indels	15;
					Gaps	2;
Qy	1013	AGGTATATTAACTATATTAAGAGAGTGTGACAAAACGATGATTTAAAGCTCGAACCGTG	1072			
Db	4539	AGGAACATCGCAACATCTTTCAGCAGCGACAGAGAAATGAAACATGAAGCCAGGATGATG	4658			



OY	1073	GCACAGCGTCAATAGTTTACGCTGGCTGGAGAGTTGAGGAGGAGAGATGCTTGAACAAG	1132
Db	4659	GCTCATGCCGGTAATCCAGCACTTTAGGAATTAGGACAGAGATCACTTGAAGCCAG	4717
OY	1133	GTTGTTCAAGGCGCAGCTGGGCAATAAACAAGATCTGTCTCTCAAAAAAAAAAAAAA	1192
Db	4718	GAGTTCAAGACCAAGCTGGGCAAATATAGTAAAGCCTTGTCTCTCAAAAAAAAAAAAAA	4776
OY	1193	AAAAAAGAAAGAGAGGGCGGGCGCTGCTGCTACGCGCTGTAATCCACGACTTTGGAG	1252
Db	4777	-----TTGGCAGGCGGGGCGCTCACGCGCTGTAATTCACGACTTTGGAG	4823
OY	1253	GCCGAGCCGGCGGATCACTGTGGTCAAGAGTTTGAACCAAGCGCTGGCCAAATGGCAA	1312
Db	4824	GCCGAGGGGCAAGTTTACTGAGGTCAAGAGTTTGAAACCAAGCGCTGGCCAAATGGCAA	4883
OY	1313	AAACCCGCTCTGTACTCAAAATGCAAAAATTTGCCAGGCGCTGTGACAGGCACTGTATCT	1372
Db	4884	AAACCCGGCTCTACTTAAAAATACAAAAATTGCTGGGGGTGGTGGCAAGCACTGTATCT	4943
OY	1373	CCAGCTACTTGGAGGCTGAGGACAGGAAGATTCGCTTGAACCAAGAGGTGAGAGTTGCAG	1432
Db	4944	CCAGCTACTTGGAGGCTGAGGACAGGAATTCATTGAATCTGGAGGTGAGAGTTGCAG	5003
OY	1433	TAACTGAGATCGTGCCTTGCACTCAAGCTTGAGGCGCAAGAGCAAGACTTGTCTCAG	1492
Db	5004	TGAGCGCGGATCGCACCTTACACTTACAGCTCGGGCAAGAGAGTGAACCTGTCTCAA	5063
OY	1493	AAAAAAAAAAAAAAAAAGAGAGAGAGAGAGAGAGAA	1531
Db	5064	AATATAATTAATAATTAATAATTAATAATTAATAA	5102

```

RESULT 15
US-10-003-295-3
; Sequence 3, Application US/10003295
; Publication No. US20020168741A1
; GENERAL INFORMATION:
; APPLICANT: GAN, weiniu et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CLO011832IV
; CURRENT APPLICATION NUMBER: US/10/003,295
; CURRENT FILING DATE: 2001-12-06
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 15297
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-003-295-3

```

Query Match	12.9%	Score 280;	DB 13;	Length 15297;
Best Local Similarity	74.9%	Pred. No. 5.7e-48;		
Matches 370;	Conservative 0;	Mismatches 110;	Indels 14;	Gaps 1;

QY	101g	ATTAACTATATAAGAGAGTTGTGATAAAAAGAGATAGATAAGTGGGAACCGTGGGACAC	1078
Db	11651	ACTCCCTTAATGCCAACCTTCCACACGAGGAGATAGATTAACCTGGCCAGTTGCTCAC	11710
QY	1079	GCTCATAGTTCTAGCTGCTGGGAGGTTGAGGAGGAGATGCGTTGAACAAGGTGTC	1138
Db	11711	GCTGTGATCCACACACTTTGGGAGCGTGAGCTGGGAGATCACTTGAAGCCAGAGGTTCC	11770
QY	1139	AAGGCGAGCTGGGCGAACATPACAGATCCTGTCTCTCAAAAAAATAAAAAAAAAAAGA	1198
Db	11771	AAGATCAGCTTGGCAACACACAGTGAATCCATCTGTACAAAAAATAACAAAAATA-----	11825
QY	1199	AAGAGAGAGGCGCGGCGTGTGGCTGACCGCTGATATCCACACACTTTGGAGGCGCCAG	1258
Db	11826	-----GACTGGGCACGGTGTCTACACTGTATATCCAGACACTTTGGGAGGCGCGAG	11876

Qy	1259	CCGGGCGGAATCACTTGGTACAGAGATTGACACCAAGCTGGCCAAATGGCAAAACCCC	1318
Db	11877	GCAGGTGAGTCACTTGTGGTCAGGAACTTTGAGACCAAGCCAGAACAAATGCTGMAACCCC	11936
Qy	1319	GTCGTGATCTCAAAATGCMAAAATTATTCACAGGCGTGTAGCAGGACCTGTAAATCCACGT	1378
Db	11937	ATCTCTACTTAAAAATTCMAAAATTTACCCAGGCACTGTGTGCAGTGTCTGTAAATCCACGT	11936
Qy	1379	ACTTGGAGGCTGAGGACAGAGAAATGCTGTGAAACCAAGAGGTGGAAGTTGCAAGTAAAGCT	1438
Db	11997	ACTTGGAGGCTGAGGCTGAGGAGAAATGCTGTGAAACCAAGAGGCGGAGGCTGCAGTGAAGCC	12056
Qy	1439	GAGATCGTGGCCGTTGCACTTCACGCCCTGGGCGACAAAGAGCAAGACTCTGTCTCAGAAAAAA	1498
Db	12057	GAGATTTGTGCCACTGCACCTGCAGCCCTGGGCGCAAGATGAAAACTTCATCTCAAAAAAAA	12116
Qy	1499	AAAAAAAAAAGAGA	1512
Db	12117	CCAAAAACAAAAA	12130

Search completed: March 13, 2004, 05:31:15  
Job time : 767.198 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: March 12, 2004, 13:19:34 ; Search time 6916.2 Seconds  
(without alignments)  
9378.080 Million cell updates/sec

Title: US-09-966-880a-15

Perfect score: 2172  
Sequence: 1 cccctgtagaggttgatga.....agtagatgataactg 2172

Scoring table: IDENTITY\_NTC  
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_estbun:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vit:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_plg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	817.4	37.6	854	13	BX464578 BX464578
2	781.2	36.0	918	13	BX391067 BX391067
3	705.2	32.5	820	13	BG755526 BG755526
4	663	30.5	664	12	BG754140 BG754140

Result No.	Score	Query Match	Length	ID	Description
5	653.2	30.1	1201	9	AL581406 AL581406
6	649	28.9	920	13	BX392040 BX392040
7	627.6	28.9	670	12	BG341546 BG341546
8	626	28.8	956	10	BF664355 BF664355
9	615.6	28.3	1201	13	BX402062 BX402062
10	572.6	26.4	604	10	AG978582 AG978582
11	567	26.1	928	12	BG398364 BG398364
12	562	25.9	570	10	AM504807 AM504807
13	498	22.9	499	13	BX283642 BX283642
14	493.6	22.7	511	12	BG170756 BG170756
15	471	21.7	517	12	BG170824 BG170824
16	467.8	21.5	528	9	AA954956 AA954956
17	466.6	21.5	939	12	BG341819 BG341819
18	457	21.0	843	12	BG758815 BG758815
19	452.8	20.8	1140	10	BF664352 BF664352
20	446.2	20.5	476	10	AW135547 AW135547
21	443	20.4	1034	12	BG755005 BG755005
22	442	20.3	442	9	AI016902 AI016902
23	441.4	20.3	464	10	AW452648 AW452648
24	406.2	18.7	422	10	AW253139 AW253139
25	369.4	17.0	374	10	AW401901 AW401901
26	362	16.7	693	10	BF975096 BF975096
27	359.2	16.5	1201	13	BX402063 BX402063
28	358.2	16.5	363	9	AI249134 AI249134
29	353.4	16.3	413	10	AW768364 AW768364
30	353	16.3	353	9	AA831307 AA831307
31	330.4	15.2	332	9	AA504653 AA504653
32	321	14.8	321	9	AI468242 AI468242
33	317.8	14.6	364	10	AW182255 AW182255
34	304.6	14.0	1201	9	AL559877 AL559877
35	280.6	12.9	581	10	AW970987 AW970987
36	280.2	12.9	730	13	BG616124 BG616124
37	279.4	12.9	4636	11	BC028737 BC028737
38	277	12.8	454	10	AW143750 AW143750
39	276.2	12.7	822	14	CB988992 CB988992
40	276.2	12.7	839	14	CD109867 CD109867
41	275.8	12.7	705	28	AQ115225 AQ115225
42	274.4	12.6	769	13	B0854207 B0854207
43	273.4	12.6	613	9	AV733538 AV733538
44	273.4	12.6	696	13	B0181548 B0181548
45	273.2	12.6	675	29	AG176894 AG176894

## ALIGNMENTS

RESULT 1  
BX464578/c 854 bp mRNA linear EST 22-MAY-2003  
LOCUS BX464578 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens CDNA  
DEFINITION clone CS00G003YB14.3-PRIME, mRNA sequence.

ACCESSION BX464578  
VERSION BX464578.1 GI:31029653

KEYWORDS EST.  
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (Bases 1 to 854)  
AUTHORS Li, W.B., Gruber, C., Jeesse, U. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 6672.r For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS10G001ZFI0NPI&cluster=6672.r. Contact :  
Feng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/Invitrogen Corporation 1600



Faraday Avenue Genoscope sequence ID : CSIDG001ZF10NP1.  
Location/Qualifiers  
Source  
1. 854

/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0D003YB14"  
/tissue\_type="B CELLS (RAMOS CELL LINE)"  
/cell\_line="RAMOS CELL LINE"  
/clone\_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"  
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

## ORIGIN

Query Match 37.6%; Score 817.4; DB 13; Length 854;  
Best Local Similarity 96.1%; Pred. No. 5.5e-110;  
Matches 816; Conservative 17; Mismatches 16; Indels 0; Gaps 0;

```

QY 307 GAACTGACAGACCTGGAGCATCTAAAGTGTCAAGTTTTCTATGACTTTAGTAGG 366
DB 854 GAACTGACAGACCTGGAGCATCTAAAGTGTCAAGTTTTCTATGACTTTAGTAGG 795
QY 367 ATGAGAGCAGAAAGTATGATCTTAAAGCATGTGAGAGATCAAAATGTTTATATCA 426
DB 794 ATGAGAGCAGAAAGTATGATCTTAAAGCATGTGAGAGATCAAAATGTTTATATCA 735
QY 427 CAACTCTTATTTATGATTCATTTAGTAAAGTGTGATGATGATTTTCTAT 486
DB 734 CAACTCTTATTTATGATTCATTTAGTAAAGTGTGATGATGATTTTCTAT 675
QY 487 CTTTCCCTTGAAGTTTATCTTCAAGTAAACAACTCTCCATCAGCCATGATCTATA 546
DB 674 CTTTCCCTTGAAGTTTATCTTCAAGTAAACAACTCTCCATCAGCCATGATCTATA 615
QY 547 GAACTCTTAATGAGATGTCGGGTGATGTGACCCCAACCAATCTCTCCAAAGCATTA 606
DB 614 GAACTCTTAATGAGATGTCGGGTGATGTGACCCCAACCAATCTCTCCAAAGCATTA 555
QY 607 ATATCCATCATGCGGTGATGTTTAAACAGACAGAGATTTTAAATGTTTATGACAA 666
DB 554 ATATCCATCATGCGGTGATGTTTAAACAGACAGAGATTTTAAATGTTTATGACAA 495
QY 667 AGAAGATGTTATGGGTGGAGATGAGAGATGACATGATGATCACTTCAAGCTACT 726
DB 494 AGAAGATGTTATGGGTGGAGATGAGAGATGACATGATGATCACTTCAAGCTACT 435
QY 727 TTAATAAGGATCTTAAATGAGGACAGAGACTGTGAAACAGACACCTTAATATGAGTT 786
DB 434 TTAATAAGGATCTTAAATGAGGACAGAGACTGTGAAACAGACACCTTAATATGAGTT 375
QY 787 GATGCTGAAGTACCAATCTTGTGAAACGCAACTCTTTTAAAGAACTCCTTAATTTA 846
DB 374 GATGCTGAAGTACCAATCTTGTGAAACGCAACTCTTTTAAAGAACTCCTTAATTTA 315
QY 847 GAAACCCCAACAACTTCAATCATTAATTAGCAACAATGGAAGAAAGTTGCTTGA 906
DB 314 GAAACCCCAACAACTTCAATCATTAATTAGCAACAATGGAAGAAAGTTGCTTGA 255
QY 907 TGTGGGAGAGAGAAATCTATGCTCTGTGGGTCTCTTCATCTCAAGAAATGCAATC 966
DB 254 TGTGGGAGAGAGAAATCTATGCTCTGTGGGTCTCTTCATCTCAAGAAATGCAATC 195
QY 967 AGGTCAAGTTTGTCAATTTGTATGTGTGATGCTTCTCCAAAGGTATATTAATA 1026
DB 194 AGGTCAAGTTTGTCAATTTGTATGTGTGATGCTTCTCCAAAGGTATATTAATA 135
QY 1027 TATAAAGAGTTGTGACAAACAGAAATGATTAAGCTGCAACCTGTGGCAACAGCTCATAG 1086
DB 134 TATAAAGAGTTGTGACAAACAGAAATGATTAAGCTGCAACCTGTGGCAACAGCTCATAG 75

```

QY 1087 TTCTAGCTCTTGGAGAGTGTGAGAGAGAGATGCTTGAACACAGGTGTTCAAGCCAG 1146  
DB 74 TTCTAGCTCTTGGAGAGTGTGAGAGAGAGATGCTTGAACACAGGTGTTCAAGCCAG 15

QY 1147 CCTGGGCA 1155  
DB 14 CCTGGGCA 6

## RESULT 2

EX391067 918 bp mRNA linear EST 13-MAY-2003  
LOCUS EX391067 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED  
DEFINITION Homo sapiens cDNA clone CS0DL012YD18 5-PRIME, mRNA sequence.  
ACCESSION EX391067  
VERSION EX391067.1 GI:30615325

## KEYWORDS

EST.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 918)  
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 6672.r For  
more information about this cluster, see

http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0BA9061ZH04\_CS05792\_1&cluster=6672.r.  
Contact : Feng Liang Email : fliang@life.techn.com URL :  
http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0BA9061ZH04\_CS05792\_1.

## FEATURES

## Source

Location/Qualifiers  
1. 918  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DL012YD18"  
/cell\_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"  
/cell\_line="RAMOS CELL LINE"  
/clone\_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

Query Match 36.0%; Score 781.2; DB 13; Length 918;  
Best Local Similarity 95.0%; Pred. No. 9.9e-105;  
Matches 872; Conservative 0; Mismatches 38; Indels 8; Gaps 6;

```

QY 134 TCTGTTTATTTCTTCACTCTCACTTCTTATAGTTTA-CAGAAAAATATTATATA- 191
DB 2 TTTTATTTTCTTAACCTTCACTTTTCTTATAGTTTACGAAAAATATTATATAC 61
QY 192 CGACTCTTAAAGATCTATGCTTGAAGAAATGAGAGACACAGCTGTGCCAGGGA 251
DB 62 CGACTCTTAAAGATCTATGCTTGAAGAAATGAGAGACACAGCTGTGCCAGGGA 121
QY 252 CGTCTGCAATTTGTGAGATTTGAATGCAACTTCCCTCTCTGGAATTAACAGACT 311
DB 122 CGTCTGCAATTTGTGAGATTTGAATGCAACTTCCCTCTCTGGAATTAACAGACT 161
QY 312 GCAGACCTGGAGACATCTTAAGTGTCAACGTTTCTATGACTTTTATAGTATGAG 371
DB 182 GCAGACCTGGAGACATCTTAAGTGTCAACGTTTCTATGACTTTTATAGTATGAG 241

```



QY 372 AGCAGAGGTAGATCTTAAAGATGAGAGATCAATGTTTATATCAATCC 431  
Db 242 AGCAGAGGTAGATCTTAAAGATGAGAGATCAATGTTTATATCAATCC 301  
QY 432 TTTATATTTGATTCATTTGAGTTAAAGTGTGTGATGATTTTCTATTCTTT 491  
Db 302 TTTATATTTGATTCATTTGAGTTAAAGTGTGTGATGATTTTCTATTCTTT 361  
QY 492 CCGTTAGAGTTTCTTCAAGTACCAAACTCTTCAAGGCGATGATTTAGGACC 551  
Db 362 CCGTTAGAGTTTCTTCAAGTACCAAACTCTTCAAGGCGATGATTTAGGACC 421  
QY 552 TCGTATGAGAGTATCGGTGATGTGACCCCAACATCTCTCCAAAGCATTAATTC 611  
Db 422 TCGTATGAGAGTATCGGTGATGTGACCCCAACATCTCTCCAAAGCATTAATTC 481  
QY 612 CAATCATGCGCTGTATGTTTAAATCAGCAGAGAGATGTTTATGTTTGAACAAAG 671  
Db 482 CAATCATGCGCTGTATGTTTAAATCAGCAGAGAGATGTTTATGTTTGAACAAAG 541  
QY 672 ATGTGATGAGTGGG--GATGAGGTATAGACCATGATGATCCTTCAAGTACTTTA 729  
Db 542 ATGTGATGAGTGGGAGTGGGAGGTATAGACCATGATGATCCTTCAAGTACTTTA 601  
QY 730 ATTAAGGATCTTAAATGAGGAGAGAGACTGTGAACAAGACCCCTAATATGAGTAT 789  
Db 602 ATTAAGGATCTTAAATGAGGAGAGAGACTGTGAACAAGACCCCTAATATGAGTAT 661  
QY 790 GTCTGAAGTACAACTCTTGTGAAAGCAGCAACTCTTTTAAAGAACTCCCTAATTAGAA 849  
Db 662 GTCTGAAGTACAACTCTTGTGAAAGCAGCAACTCTTTTAAAGAACTCCCTAATTAGAA 721  
QY 850 ACAACCAACAACTTCAATCATTAATTAAGCAAAACATTTG--AAGAAAGTCTTGATG 908  
Db 722 ACAACCAACAACTTCAATCATTAATTAAGCAAAACATTTG--AAGAAAGTCTTGATG 781  
QY 909 TTGGGAGAGAGAAATCTATGTGCTGCTGCTGCTCTTCAATCTCAAGAAATGCAATCG 968  
Db 782 TTGGGAGAGAGAAATCTATGTGCTGCTGCTGCTCTTCAATCTCAAGAAATGCAATCG 839  
QY 969 GTCAAGGTTTGTACATTTTGTATGTGTGATGCTTCCCAAGGTAATTAATATA 1028  
Db 840 GTCAAGGTTTGTACATTTTGTATGTGTGATGCTTCCCAAGGTAATTAATATA 898  
QY 1029 TAAGAGTTGTGACAAA 1046  
Db 899 TAAGAGTTGTGACACA 916

RESULT 3  
BG755526 820 bp mRNA linear EST 15-MAY-2001  
LOCUS 602716206F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:485656 5',  
DEFINITION mRNA sequence.  
ACCESSION BG755526  
VERSION BG755526.1 GI:14066179  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 820)  
NIH-MGC http://mgs.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strauberg, Ph.D.  
Email: csaps-r@mail.nih.gov  
Tissue Procurement: Louis M. Straub, M.D., Ph.D.  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incey Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:

FEATURES  
SOURCE  
http://image.lnl.gov  
Plate: LNCM1707 row: k column: 01  
High quality sequence stop: 638.  
Location/Qualifiers  
1..820  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4856568"  
/tissue="primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_id="NIH\_MGC\_48"  
/note="Organ: B-cells; Vector: pOTB7; Site: 1; XhoI;  
Site 2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGACGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by Ling  
Hong in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 32.5%; Score 705.2; DB 12; Length 820;  
Best Local Similarity 94.5%; Pred. No. 1.3e-93;  
Matches 775; Conservative 0; Mismatches 38; Indels 7; Gaps 4;

QY 12 GGTGATGACTTACAGAGCATTCGTAATTTGGAGCTTGTATGCAATTCAGGAAT 71  
Db 2 GGTGATGACTTACAGAGCATTCGTAATTTGGAGCTTGTATGCAATTCAGGAAT 60  
QY 72 GTCAACACGATGAAATCTCTGCTGAACAGGTGATTAACAAACGCTTCAAGTCT 131  
Db 61 GTCAACACGATGAAATCTCTGCTGAACAGGTGATTAACAAACGCTTCAAGTCT 120  
QY 132 TCTGCTTTTATCTTCAACTCTCACTTCTTGAAGTATTAAGAAAAATATTATATA 191  
Db 121 TCTGCTTTTATCTTCAACTCTCACTTCTTGAAGTATTAAGAAAAATATTATATA 180  
QY 192 CGACTCTTAAAGATCTATGCTTGAATAAAGAGAGAACACAGCTGSCCAGGA 251  
Db 181 CGACTCTTAAAGATCTATGCTTGAATAAAGAGAGAACACAGCTGSCCAGGA 240  
QY 252 CGTGTGCAATTTGTGAGTTTGAATGCAATTTGCCCTACTGGAATTAACAGAACT 311  
Db 241 CGTGTGCAATTTGTGAGTTTGAATGCAATTTGCCCTACTGGAATTAACAGAACT 300  
QY 312 GCAGAGCCTGGAGACATCTTAAGTGTCAAGCTTTTCTATGACTTTAGTAGAGATG 371  
Db 301 GCAGAGCCTGGAGACATCTTAAGTGTCAAGCTTTTCTATGACTTTAGTAGAGATG 360  
QY 372 AGCAGAGGTAGATCTTAAAGATGAGAGATCAATGTTTATATCAATCC 431  
Db 361 AGCAGAGGTAGATCTTAAAGATGAGAGATCAATGTTTATATCAATCC 420  
QY 432 TTTATATTTGATTCATTTGAGTTAAAGTGTGTGATGATTTTCTATTCTTT 491  
Db 421 TTTATATTTGATTCATTTGAGTTAAAGTGTGTGATGATTTTCTATTCTTT 480  
QY 492 CCGTTAGAGTTTCTTCAAGTACCAAACTCTTCAAGGCGATGATTTAGGACC 551  
Db 481 CCGTTAGAGTTTCTTCAAGTACCAAACTCTTCAAGGCGATGATTTAGGACC 540  
QY 552 TCGTATGAGAGTATCGGTGATGTGACCCCAACATCTCTCCAAAGCATTAATTC 611  
Db 541 TCGTATGAGAGTATCGGTGATGTGACCCCAACATCTCTCCAAAGCATTAATTC 600  
QY 612 CAATCATGCGCTGTATGTTTAAATCAGCAGAGAGATGTTTATGTTTGAACAAAG 671  
Db 601 CAATCATGCGCTGTATGTTTAAATCAGCAGAGAGATGTTTATGTTTGAACAAAG 660  
QY 672 ATGTGATGAGTGGGAGTATGAGGTAT--AGACCATGATGATCCTTCAAG--CTACT 726



Db 661 ATTGTCATGGTGGGATGCGCGCTCTCAGACATGATGGTCACTTCAAGCTACTT 720  
Qy 727 TTAATTAAGGATCTTAAATGGGACAGAGACTGTGGAACAAGACCCCTAATAGGTT 786  
Db 721 TTAATTAAGGATCTCAGATGGGACAGAGACTGTGGAACAAGAAATCTTAATATGGGT 780  
Qy 787 GATGCTGAAG-TAGCAATCTTCTGGAAACGCAACTCT 825  
Db 781 GATGCTGAAGTTAGCAATCTCTGGAAACGCAACTCT 820

RESULT 4  
Bg754140 664 bp mRNA linear EST 15-MAY-2001  
LOCUS 602709681F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4846005 5',  
DEFINITION mRNA sequence.  
ACCESSION Bg754140  
VERSION Bg754140.1 GI:14064793  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE NIH-MGC http://mgi.nci.nih.gov/  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE Unpublished (1999)  
JOURNAL Contact: Robert Strausberg, Ph.D.  
COMMENT Email: csapsb-remail.nih.gov  
Tissue Procurement: Louis M. Straud, M.D., Ph.D.  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.lnl.gov  
Plate: LNCM686 row: b column: 22  
High quality sequence stop: 662.  
Location/Qualifiers  
1. 664  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4846005"  
/issue\_type="primary B-cells from tonsils (cell line)"  
/lab host="DH10B (phage-resistant)"  
/clone\_1lb="NIH\_MGC\_48"  
/note="Organ: B-cells; Vector: pOT7; Site 1: XhoI;  
Site 2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCAAGG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by Ling  
Hong in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."

ORIGIN  
Query Match 30.5%; Score 663; DB 12; Length 664;  
Best Local Similarity 100.0%; Prid. No. 2.1e-87;  
Matches 663; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 GGTGATGACTTACGAGACGATTCGTACTTGGGACTTGTAGCACTTCAGAAAT 71  
Db 2 GGTGATGACTTACGAGACGATTCGTACTTGGGACTTGTAGCACTTCAGAAAT 61  
Qy 72 GTCAACACGATGAATATCTCTGCTGAGACAGTGATGATTAATAAAGAGCTCTCAAGTCT 131  
Db 62 GTCAACACGATGAATATCTCTGCTGAGACAGTGATGATTAATAAAGAGCTCTCAAGTCT 121  
Qy 132 TCTGTGTTTATCTTCAACTCTCACTTCTTGAAGTTTACAGAAAAATATTATTA 191  
Db 122 TCTGTGTTTATCTTCAACTCTCACTTCTTGAAGTTTACAGAAAAATATTATTA 181

Qy 192 CGACTCTTAAAAAGATCTATGCTTGAATAAGAGAGAAACAAGGCTGGGCGAGGA 251  
Db 182 CGACTCTTAAAAAGATCTATGCTTGAATAAGAGAGAAACAAGGCTGGGCGAGGA 241  
Qy 252 CGTGTGCAATGTGTCAGCTTTGAATGCAATGTCCCTTACTGGGAATTAACAAGT 311  
Db 242 CGTGTGCAATGTGTCAGCTTTGAATGCAATGTCCCTTACTGGGAATTAACAAGT 301  
Qy 312 GCAAGACTGGGAGCATCTTAAGTGTCAAGCTTTTCTATGACTTTAGTGGATGAG 371  
Db 302 GCAAGACTGGGAGCATCTTAAGTGTCAAGCTTTTCTATGACTTTAGTGGATGAG 361  
Qy 372 AGCAGAACTGATGCTTAAATAAGATGTGAGAGATCAATGTTTATCAACATCC 431  
Db 362 AGCAGAACTGATGCTTAAATAAGATGTGAGAGATCAATGTTTATCAACATCC 421  
Qy 432 TTTATTTATGATTCATTTGATTAACAGTGTGTTAGTATGATTTTCTATTCTTT 491  
Db 422 TTTATTTATGATTCATTTGATTAACAGTGTGTTAGTATGATTTTCTATTCTTT 481  
Qy 492 CCTTGACGTTTACTTCAAGTACACAACTCTTCCATCAGGCCATGATTAAGACC 551  
Db 482 CCTTGACGTTTACTTCAAGTACACAACTCTTCCATCAGGCCATGATTAAGACC 541  
Qy 552 TCCCTAATGAGATGATCGGGTGTGATGAGACCCCAACATCTCTCAAGCATTAATATC 611  
Db 542 TCCCTAATGAGATGATCGGGTGTGATGAGACCCCAACATCTCTCAAGCATTAATATC 601  
Qy 612 CAATCATGCGCTGATGTTTAAATCAGCAGAACATGTTTATGTTGTGACAAAAGAG 671  
Db 602 CAATCATGCGCTGATGTTTAAATCAGCAGAACATGTTTATGTTGTGACAAAAGAG 661  
Qy 672 ATT 674  
Db 662 ATT 664

RESULT 5  
AL581406/c 1201 bp mRNA linear EST 01-JUN-2003  
LOCUS AL581406 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA  
DEFINITION clone CS0DG003YB14 3-PRIME, mRNA sequence.  
ACCESSION AL581406  
VERSION AL581406.2 GI:31319646  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE Li, W.B., Gruber, C., Jesse, J., and Polayes, D.  
1 (bases 1 to 1201)  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT On Feb 16, 2001 this sequence version replaced gi:12948375.  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
was not normalized. Library was constructed by Life Technologies, a  
division of Invitrogen. This sequence belongs to sequence cluster  
6672.r For more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0DG003DA07NP1&cluster=6672.r. Contact :  
Feng Liang Email : fliang@life.techn.com URL :  
http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
Paradise Avenue Genoscope sequence ID : CS0DG003DA07NP1.  
Location/Qualifiers  
1. 1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DG003YB14"



/cissue\_type="B CELLS (RAMOS CELL LINE)"  
 /cell\_line="RAMOS CELL LINE"  
 /clone\_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"  
 /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed  
 with a NotI-oligo(dT) primer. Five prime end enriched,  
 double-strand cDNA was digested with Not I and cloned into  
 the Not I and EcoRV sites of the pCMVSPORT 6 vector.  
 Library was not normalized."

Query Match 30.1%; Score 653.2; DB 9; Length 1201;  
 Best Local Similarity 89.8%; Pred. No. 3.8e-86;  
 Matches 680; Conservative 31; Mismatches 42; Indels 4; Gaps 2;

QY 404 AGAGTAAGTATTTTATATCAACATCTTTATTTATGATTCATTGATTAAGAGG 463  
 DB 756 AADATATCAAAATTTTATATCAAAATCTTTATTTATTTATTCATTATTA-TTAACATTGT 698  
 QY 464 TGTATGATAGATTTTCTATCTTTCCCTGAGTTTACTTCAAGTAAACAACACT 523  
 DB 697 TTMTATWDATATATTTTCTATTTTCTCTGKSGTTKACTACAAATAGAGAAATT 638  
 QY 524 CTTCATCAAGGCATGATCTATAGACCTCTTAATGAGATCTGGGTGATTGTGACC 583  
 DB 637 CTTSSTTSAKGCATATCTATAGGGCCCTTAATGAGAGTANCTGGGTGATTGTGACC 578  
 QY 584 CAACCACTCTTCCAAAGCATTTATTCATCATGCGCTGATGTTTATCAGCAGAA 643  
 DB 577 CAACCAACNCTCCAAAGCATTTATTCATCATGCGCTGATGTTTAAACACAGAA 518  
 QY 644 GCATGTTTATGTTTGTATCAAAAGAGTGTATGAGTGGGATGAGAGTATGACCA 703  
 DB 517 GCATGTTTATGTTTGTATCAAAAGAGTGTATGAGTGGGATGAGAGTATGACCA 458  
 QY 704 TGATGATGATCCTTCAAGCTCTTATTAAGAGTCTTAAATGGGAGAGAGAGTGTGA 763  
 DB 457 TGATGATGATCCTTCAAGCTCTTATTAAGAGTCTTAAATGGGAGAGAGAGTGTGA 398  
 QY 764 ACAAGACACCTTATTAATGAGTGTATGATGATGAGTATCTTCTGGAACGCAACT 823  
 DB 397 ACAAGACACCTTATTAATGAGTGTATGATGATGAGTATCTTCTGGAACGCAACT 338  
 QY 824 CTTTAAAGAGTCCCTTAATTTAGAAACCCCAACACTTCAATATCATATTTGCAAA 883  
 DB 337 CTTTAAAGAGTCCCTTAATTTAGAAACCCCAACACTTCAATATCATATTTGCAAA 278  
 QY 884 CAATTGAGAGAGTGTGATGATGTTGGGAGAGAGAAATCTATTGCTCGTGGGTC 943  
 DB 277 CAATTGAGAGAGTGTGATGATGTTGGGAGAGAGAAATCTATTGCTCGTGGGTC 218  
 QY 944 TCTTCATCTCAGAAATGCAATCAGGTCAAGGTTTGCTCATTTTGTATGTGATGTC 1003  
 DB 217 TCTTCATCTCAGAAATGCAATCAGGTCAAGGTTTGCTCATTTTGTATGTGATGTC 158  
 QY 1004 TTCTCCCAAGGTATATTAATATATATAGAGTGTGACAAACAGAAATGATAAGCTG 1063  
 DB 157 TTCTCCCAAGGTATATTAATATATATAGAGTGTGACAAACAGAAATGATAAGCTG 98  
 QY 1064 CGAACCGTGGACACCTCATAGTTTACGTTGGGAGGTTGAGAGAGAGAGTGGCT 1123  
 DB 97 CGAACCGTGGACACCTCATAGTTTACGTTGGGAGGTTGAGAGAGAGAGTGGCT 38  
 QY 1124 TGAACACAGGTGTTCAAGGCGACGCTGGGCAACATTA 1160  
 DB 37 ---GAAACACAGTTCAGGCGACGCTGGGCAACATTA 4

RESULT 6  
 BX392040/c 920 bp mRNA linear EST 13-MAY-2003  
 LOCUS BX392040 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED  
 DEFINITION Homo sapiens CDNA clone CS0DL012YD18 3-PRIME, mRNA sequence.  
 ACCESSION BX392040

VERSION BX392040.1 GI:30607824  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 920)  
 AUTHORS Li, W. B., Gruber, C., Jesse, J., and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Library was constructed by life technologies, a division of  
 invitrogen. This sequence belongs to sequence cluster 6672.r For  
 more information about this cluster, see  
 http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CS0BA1048ZH08\_CS04576\_1cluster=6672.r.  
 Contact : Peng Liang Email : filiang@lifetech.com URL :  
 http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID : CS0BA1048ZH08\_CS04576\_1.  
 Location/Qualifiers  
 1..920  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0DL012YD18"  
 /cell\_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"  
 /clone\_lib="RAMOS CELL LINE"  
 /clone\_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT  
 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo(dT)  
 primer. Five prime end enriched, double-strand cDNA was  
 digested with Not I and cloned into the Not I and EcoRV  
 sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

Query Match 29.9%; Score 649; DB 13; Length 920;  
 Best Local Similarity 98.5%; Pred. No. 1.8e-85;  
 Matches 655; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 CCCTGTATAGAGTGTATGATCTTACGAGAGCACTTTCGTAATCTTGGACCTTGAATGCA 60  
 DB 665 CCCTGTATAGAGTGTATGATCTTACGAGAGCACTTTCGTAATCTTGGACCTTGAATGCA 606  
 QY 61 CTTCCAGAGATGTCACACAGATGAATATCTGCTGAGAGAGAGATGATTAATAACAGT 120  
 DB 605 CTTCCAGAGATGTCACACAGATGAATATCTGCTGAGAGAGAGATGATTAATAACAGT 546  
 QY 121 CTTCAAGTCTTCTGTTTATTTATTTCAACTCTCACTTCTTAGAGTTTACAGAAAA 180  
 DB 545 CTTCAAGTCTTCTGTTTATTTATTTCAACTCTCACTTCTTAGAGTTTACAGAAAA 486  
 QY 181 ATATTATATACGACTTTTAAAGAGTATGTTCTTGAAGATGAGAGAGAAACACAGT 240  
 DB 485 ATATTATATACGACTTTTAAAGAGTATGTTCTTGAAGATGAGAGAGAAACACAGT 426  
 QY 241 CTGGCCAGGAGACGTGCTGCAATTTGGGCAATTTTGAATGGAATGTCCTCCCACTGGGA 300  
 DB 425 CTGGCCAGGAGACGTGCTGCAATTTGGGCAATTTTGAATGGAATGTCCTCCCACTGGGA 366  
 QY 301 ATAAACAGATGTCAGACCTGGAGGATCCTTAAGTGTCAACGTTTCTATGACTTTTA 360  
 DB 365 ATAAACAGATGTCAGACCTGGAGGATCCTTAAGTGTCAACGTTTCTATGACTTTTA 306  
 QY 361 GGTAGATGAGAGAGAGTATGATCTTAAAGAGTGTGAGAGATCAATGTTTAA 420  
 DB 305 GGTAGATGAGAGAGAGTATGATCTTAAAGAGTGTGAGAGATCAATGTTTAA 246  
 QY 421 TATCAACATCTTATTTATTTGATTCATTTGAGTTAACAGTGTGTTAGTAGATTTT 480  
 DB 245 TATCAACATCTTATTTATTTGATTCATTTGAGTTAACAGTGTGTTAGTAGATTTT 186



QY 481 TCTATCTTCCCTGACGTTAATTCAAGTAACACAACTCTTCATCAGGCGATGA 540  
DB 185 TCTATCTTCCCTGACGTTAATTCAAGTAACACAACTCTTCATCAGGCGATGA 126  
QY 541 TCTATGAGACCTCTTATGAGAGATCTGGGTGATTGTGACCCCAACATCTCTCCAA 600  
DB 125 TCTATGAGACCTCTTATGAGAGATCTGGGTGATTGTGACCCCAACATCTCTCCAA 66  
QY 601 GCATTATATCCATATGCGCTGATGTTTAAATCAGACAGAGATGTTTATGTTG 660  
DB 65 GATTATATATCAATCATGCGCTGATGTTTAAATCAGACAGAGATGTTTATGTTG 6  
QY 661 TACA 665  
DB 5 AAAAA 1

RESULT 7  
LOCUS BG341546 670 bp mRNA linear EST 27-FEB-2001  
DEFINITION 602463652P1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4576548 5',  
mRNA sequence.  
ACCESSION BG341546  
VERSION BG341546.1 GI:13147984  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 670)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
COMMENT Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-remail.nih.gov  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
Plate: LNCM1289 row: g column: 13  
High quality sequence stop: 650.  
Location/Qualifiers  
1. 670  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4576548"  
/rissue\_type="primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_1lb="NIH\_MGC\_48"  
/note="Organ: B-cells; Vector: pOT8; Site\_1: XhoI;  
Site\_2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGACGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by Ling  
Hong in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC library."

ORIGIN  
Query Match 28.8%; Score 627.6; DB 12; Length 670;  
Best Local Similarity 98.8%; Pred. No. 3e-82;  
Matches 664; Conservative 0; Mismatches 4; Indels 4; Gaps 3;

QY 12 GATTGATGACTTACGAGACGATTTGTACTTTGGGACTTGTAGCACTTCAGGAAT 71  
DB 2 GATTGATGACTTACGAGACGATTTGTACTTTGGGACTTGTAGCACTTCAGGAAT 60

QY 72 GTCAACACGATGAATAATCTCTGCTGAAGACAGTGTGATAAAACAGTCTTCAAGTCT 131  
DB 61 GTCAACACGATGAATAATCTCTGCTGAAGACAGTGTGATAAAACAGTCTTCAAGTCT 120  
QY 132 TCTCTGTTTATCTTCAACTCTCAGCTTCTTTAGGTTTACAGAAAAATTTATATA 191  
DB 121 TCTCTGTTTATCTTCAACTCTCAGCTTCTTTAGGTTTACAGAAAAATTTATATA 180  
QY 192 CGACTCTTTAAAGATCTATGCTTGAAGAAATGAGAGAAACAGAGTCTGGCCAGGGA 251  
DB 181 CGACTCTTTAAAGATCTATGCTTGAAGAAATGAGAGAAACAGAGTCTGGCCAGGGA 240  
QY 252 CGTGTGCAATTTGTGAGTTTGAATGCAATGTCCCTTACTGGGAATTAACAAGTCT 311  
DB 241 CGTGTGCAATTTGTGAGTTTGAATGCAATGTCCCTTACTGGGAATTAACAAGTCT 300  
QY 312 GCAGGACCTGGAGCATCTTAAAGTGTCAACGTTTCTATAGCTTTAGTAGATGAG 371  
DB 301 GCAGGACCTGGAGCATCTTAAAGTGTCAACGTTTCTATAGCTTTAGTAGATGAG 360  
QY 372 AGCAGAAAGTAGATCTTAAAGATGTGAGAGATCAAAATGTTTATATCAATCC 431  
DB 361 AGCAGAAAGTAGATCTTAAAGATGTGAGAGATCAAAATGTTTATATCAATCC 420  
QY 432 TTTATTTATTTGATTCATTTGAGTTTAAAGTGTGATGATGATTTTCTATCTTT 491  
DB 421 TTTATTTATTTGATTCATTTGAGTTTAAAGTGTGATGATGATTTTCTATCTTT 480  
QY 492 CCCTTGACGTTTACTTGAAGTAAACAACTCTTCATGAGCCATGATCTATAGACC 551  
DB 481 CCCTTGACGTTTACTTGAAGTAAACAACTCTTCATGAGCCATGATCTATAGACC 540  
QY 552 TCCCTAAGAGATATCTGGGTGATTTGGACCCCAACATCTCTCC-AAAGCTATATAT 610  
DB 541 TCCCTAAGAGATATCTGGGTGATTTGGACCCCAACATCTCTCCAAAGCTATATAT 600  
QY 611 CCATCATGCGCTGATGTTTATATCAGCAGAAAGATGTTTATGTTTATACAAAGAA 670  
DB 601 CCATCATGCGCTGATGTTTATATCAGCAGAAAGATGTTT--AGTTGTACCAAGAA 658  
QY 671 GATTGTTATGGG 682  
DB 659 GATTGTTATGGG 670

RESULT 8  
LOCUS BF664355 956 bp mRNA linear EST 21-DEC-2000  
DEFINITION 602146013P1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4309707 5',  
mRNA sequence.  
ACCESSION BF664355  
VERSION BF664355.1 GI:11938351  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 956)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
COMMENT Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-remail.nih.gov  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
Plate: LNCM1183 row: i column: 04  
High quality sequence stop: 725.  
Location/Qualifiers

FEATURES



## Source

1. .956  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAG:4309707"  
 /issue\_type="primary B-cells from consils (cell line)"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="XIH MGC 48"  
 /note="Organ: B-cells; Vector: pOTB1; Site 1: XhoI;  
 Site 2: EcoRI; cDNA made by oligo-dT priming.  
 Directionally cloned into EcoRI/XhoI sites using the  
 following 5' adaptor: GGCAAGAG(G). Size-selected >500bp  
 for average insert size 1.8kb. Library constructed by Ling  
 Hong in the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies).  
 Note: this is a NIH\_MGC library."

## ORIGIN

Query Match 28.8%; Score 626; DB 10; Length 956;  
 Best Local Similarity 97.2%; Pred. No. 4.1e-82;  
 Matches 690; Conservative 0; Mismatches 15; Indels 5; Gaps 5;

QY 12 GGTGATGACTTACGAGACGATTTGCTACTTTGGACTTTGATAGCACTTCCAGGAT 71  
 DB 2 GGTGATGACTTACGAGACGATTTGCTACTTTGGACTTTGATAGCACTTCCAGGAT 60  
 QY 72 GTCAACACGATGAATATCTCTGCTGAAGACAGTGATTAATAACAGTCTTCAA-GTC 130  
 DB 61 GTCAACACGATGAATATCTCTGCTGAAGACAGTGATTAATAACAGTCTTCAAAGTTC 120  
 QY 131 TTCTCTGTTTATCTTCAACTCTGCTTCTTAAGTTTACAGAAAAATTTATAT 190  
 DB 121 TTCTCTGTTTATCTTCAACTCTGCTTCTTAAGTTTACAGAAAAATTTATAT 180  
 QY 191 ACCAGCTTTAAAGATGCTATGCTTGAAGATAGAGAAACAGAGTCTGGCAGGG 250  
 DB 181 ACCAGCTTTAAAGATGCTATGCTTGAAGATAGAGAAACAGAGTCTGGCAGGG 240  
 QY 251 ACCGTGTCATTTGTGAGTTTGAATGCAATTTGCCCTACTGGAGATTAAGAGAC 310  
 DB 241 ACCGTGTCATTTGTGAGTTTGAATGCAATTTGCCCTACTGGAGATTAAGAGAC 300  
 QY 311 TGCAGACCTGGAGACATCTTAAGTCAAGCTTTTCTATGACTTTAGGTAGAGATCA 370  
 DB 301 TGCAGACCTGGAGACATCTTAAGTCAAGCTTTTCTATGACTTTAGGTAGAGATCA 360  
 QY 371 GACAGAGGTAGATCTTAAGATGAGATGAGATCAATTTTATATCAACATC 430  
 DB 361 GACAGAGGTAGATCTTAAGATGAGATGAGATCAATTTTATATCAACATC 420  
 QY 431 CTTATATTTAGTATCTT-TGAGTTTAAAGTGTGATAGATTTTCTATTTCT 489  
 DB 421 CTTATATTTAGTATCTT-TGAGTTTAAAGTGTGATAGATTTTCTATTTCT 480  
 QY 490 TTCCCTGACGTTTACTTCAAGTAAACAACCTTTCATCAAGCCATGATCTATAGGA 549  
 DB 481 TTCCCTGACGTTTACTTCAAGTAAACAACCTTTCATCAAGCCATGATCTATAGGA 540  
 QY 550 CTTCTTAATGAGAGATCTGGGATGATGAGACCCCAACCATCTTCCCAAGCATTAATA 609  
 DB 541 CTTCTTAATGAGAGATCTGGGATGATGAGACCCCAACCATCTTCCCAAGCATTAATA 599  
 QY 610 TCCAAATGAGAGATGATGTTTAAATCAGAGAGAGATTTTATGTTTATCAAAAGA 669  
 DB 600 TCCAAATGAGAGATGATGTTTAAATCAGAGAGAGATTTTATG-TTGTACAAAGA 658  
 QY 670 AGATTGTATGGGTGGGATGAGAGATAGACCATGATGATCACTTCA 719  
 DB 659 AGATTGTATGGGTGGGATGAGAGATAGACCATGATGATCACTTCA 708

RESULT 9

EX402062/c  
 LOCUS 1201 bp mRNA linear EST 13-MAY-2003  
 DEFINITION EX402062 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED  
 Homo sapiens cDNA clone CS0D012BD18 3-PRIME, mRNA sequence.  
 ACCESSION EX402062  
 VERSION EX402062.1 GI:30626643  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 1201)  
 Li, W.-B., Gruber, C., Jesse, J. and Polayes, D.  
 Full-length cDNA libraries and normalization  
 Unpublished (2001)  
 JOURNAL CONTACT: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 6672.r For  
 more information about this cluster, see  
 http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CS0D012BD09NP1&cluster=6672.r. Contact :  
 Feng Liang Email : fliang@lifetech.com URL :  
 http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID : CS0D012BD09NP1.  
 Location/Qualifiers

## FEATURES

source  
 1. .1201  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CS0D012BD18"  
 /cell\_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"  
 /cell\_line="RAMOS CELL LINE"  
 /clone\_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT  
 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo(dT)  
 primer. Five prime end enriched, double-strand cDNA was  
 digested with Not I and cloned into the Not I and EcoR V  
 sites of the pCMVSPORT 6 vector. Library was normalized."

## ORIGIN

Query Match 28.3%; Score 615.6; DB 13; Length 1201;  
 Best Local Similarity 77.0%; Pred. No. 1.1e-80;  
 Matches 715; Conservative 92; Mismatches 107; Indels 15; Gaps 8;

QY 204 AAGATCTATGCTTGAAGATTAAGAGAGACACAGGCTGCCAGGAGCGTGCAAT 263  
 DB 945 AAAAAATWASWAKGTTPAATRTDRAKAGACASAGTGYKGGKSGSKCAAT- 887  
 QY 264 GGTGACGTTTGAATGACATTTGTCCTTACTGGATTAACAGACTGACGAGCCTGG 323  
 DB 886 -GKTGCKRTGTGAKGACATTTGTCM-AAAAAAAAAAAAAAAAANRR-TMNNMT 829  
 QY 324 AGCATCTTAAGTGTCAAGCTTTTCTATGACTTTTNGTGAAGATGAGACGAGATG 383  
 DB 828 AKNNWCCCTAAAGTCAACCTTATCTATGACAAWAAAGTAGATKAKACAAAGATAD 769  
 QY 384 ATCTTAAGAGATGAGATGATCAAGTTTATATCAACATCTTATATTTATTTGA 443  
 DB 768 ATCTTAAGAGATGAGATGATCAAGTTTATATCAACATCTTATATTTATTTGA 709  
 QY 444 TTCAATTTAGTACAGTGGGT-TAGTGAATGTTTCTTATCTTCTTCTTGAAGT 502  
 DB 708 TWCAATTTAKATWACATGKTGTATAGATATGATTTTCTATTTTAAACCTTGAAGT 649  
 QY 503 T---ACTTCAAGTACACAACTCTTCACTAGGCCATGATCTATAGACCTTCAATG 559  
 DB 648 TCACNCTACACAGTAAACAACCTTCAACAGGCATGATCTATAGACCTCCCAAG 589  
 QY 560 AGATTATGGGTATTTGACCCCAACCATCTTCCAAAGCATTAATCCAAATATG 619







## FEATURES

Location/Qualifiers  
1. .928  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:456366"  
/tissue\_type="primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_id="NH\_MGC\_48"  
/note="Organ: B-cells; Vector: pOTB; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

## ORIGIN

Query Match 26.1%; Score 567; DB 12; Length 928;  
Best Local Similarity 98.6%; Pred. No. 1.7e-73;  
Matches 614; Conservative 0; Mismatches 5; Indels 4; Gaps 4;

```

QY 1 CCCCTGATGAGGTGATGACCTACGACGACATTCGACTTGGGACCTTGATGCA 60
DB 56 CCCCTGATGAGGTGATGACCTACGACGACATTCGACTTGGGACCTTGATGCA 114
QY 61 CTCACGAGATGTCACACAGATGAATATCTCTGCTGAGACAGTGAATAAACAGT 120
DB 115 CTCACGAGATGTCACACAGATGAATATCTCTGCTGAGACAGTGAATAAACAGT 174
QY 121 CCTCAAGTCTCTCTGTTTATTTCTTCACTCTCACTTCTTGAAGTTACAGAAAA 180
DB 175 CCTCAAGTCTCTCTGTTTATTTCTTCACTCTCACTTCTTGAAGTTACAGAAAA 234
QY 181 ATATTTATATGACCTTTTAAAGATCTATGCTTGAATAGAAAGAAACAGAGT 240
DB 235 ATATTTATATGACCTTTTAAAGATCTATGCTTGAATAGAAAGAAACAGAGT 294
QY 241 CTGGCCGAGGACGCTGCAATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 300
DB 295 CTGGCCGAGGACGCTGCAATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 353
QY 301 ATACAGACGACGACGCTGGAGCATCTTAAGTGTCAAGCTTTTCTATGACTTTA 360
DB 354 ATACAGACGACGACGCTGGAGCATCTTAAGTGTCAAGCTTTTCTATGACTTTA 413
QY 361 GGATGATGAGACGAGAGTGAATCTTAAAGATGGTGAAGAGATCAATG-TTTT 419
DB 414 GGATGATGAGACGAGAGTGAATCTTAAAGATGGTGAAGAGATCAATGTTT 473
QY 420 AATACACATCTTTATTTATTTGATTTGATTTGATTTGATTTGATTTGATTT 479
DB 474 AATACACATCTTTATTTATTTGATTTGATTTGATTTGATTTGATTTGATTT 533
QY 480 TTCAATCTTTCCCTGAGCTTACTTTCAAGTAAACAACTCTTCCATCAGGCATG 539
DB 534 TTCAATCTTTCCCTGAGCTTACTTTCAAGTAAACAACTCTTCCATCAGGCATG 593
QY 540 ATCTATGACCTCTTATAGAGTATCTGGTGGTGGTGGTGGTGGTGGTGGTGGT 599
DB 594 ATCTATGACCTCTTATAGAGTATCTGGTGGTGGTGGTGGTGGTGGTGGTGGT 652
QY 600 AGCATTAATATCCATCATGCGC 622
DB 653 AGCATTAATATCCATCATGCGC 675

```

RESULT 12  
AM504807 570 bp mRNA linear EST 02-MAR-2000  
LOCUS  
DEFINITION UI-HF-BNO-ain-f-05-0-UI.t1 NIH\_MGC\_50 Homo sapiens cDNA clone

IMAGE:3080216 5', mRNA sequence.

ACCESSION  
AM504807  
VERSION  
AM504807.1 GI:7142474  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE  
NIH-MGC <http://mgs.nci.nih.gov/>.  
TITLE  
Unpublished (1999)  
JOURNAL  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Eco RI site shown at the beginning of the sequence.  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
CDNA Library Preparation: M.B. Soares Lab  
CDNA Library Arrayed by: M.B. Soares Lab  
DNA Sequencing by: M.B. Soares Lab  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at:  
[www.bio.linnl.gov/bdip/image/image.html](http://www.bio.linnl.gov/bdip/image/image.html)  
Seq primer: M13 Forward.

## FEATURES

Location/Qualifiers  
1. .570  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3080216"  
/tissue\_type="lymph"  
/cell\_type="germinal center B cells"  
/cell\_line="MGC85"  
/lab\_host="DH10B (LTI)"  
/clone\_id="NH\_MGC\_50"  
/note="Vector: pTZ19-Pac; Site 1: NotI; Site 2: Eco RI; Constructed from size fractionated cytoplasmic mRNA (3.5-4.4kb). Directionally cloned. Cells provided by Louis M. Staudt, Ph.D. Library preparation by Maria de Fatima Bonaldo, Ph.D. and M. Bento Soares, Ph.D."

## ORIGIN

Query Match 25.9%; Score 562; DB 10; Length 570;  
Best Local Similarity 99.8%; Pred. No. 1.2e-72;  
Matches 562; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 489 TTTCCTTGAAGTTTCACTTCAAGTAAACAACTCTTCCATCAGGCATGATATAGG 548
DB 8 TTTCCTTGAAGTTTCACTTCAAGTAAACAACTCTTCCATCAGGCATGATATAGG 67
QY 549 ACCTCTAATGAGATCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 608
DB 68 ACCTCTAATGAGATCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 127
QY 609 ATCCATATGCGCTGATGTTTATATCAGAGAAAGATGTTTATGTTTGTCAAAAG 666
DB 128 ATCCATATGCGCTGATGTTTATATCAGAGAAAGATGTTTATGTTTGTCAAAAG 187
QY 669 AAGATGTTATGGGAGGAGTGAAGTATGACCATGATGGTCACTTCAAGCTACTTT 728
DB 188 AAGATGTTATGGGAGGAGTGAAGTATGACCATGATGGTCACTTCAAGCTACTTT 247
QY 729 AATAAGATCTTAAATGGGACAGAGACTGTGAACAAGACACCTAATATATGGTTGA 788
DB 248 AATAAGATCTTAAATGGGACAGAGACTGTGAACAAGACACCTAATATATGGTTGA 307
QY 789 TGTCTGAAGTACCAATTTCTGGAACGCAATCTTTTAAAGAGTCCCTAATTTTGA 848
DB 308 TGTCTGAAGTACCAATTTCTGGAACGCAATCTTTTAAAGAGTCCCTAATTTTGA 367
QY 849 AACCCCAACAACCTTCATATCATTAATTACCAACAATTGGAAGAGTGGCTTGAATG 908
DB 368 AACCCCAACAACCTTCATATCATTAATTACCAACAATTGGAAGAGTGGCTTGAATG 427

```



QY	909	TTGGGAGAGAGAAATCTATTTGGCTTCGTCGGTCTCTTCAATCTCAGAAATGCCAATCAG	968
Db	428	TTGGGAGAGAGAAATCTATTTGGCTTCGTCGGTCTCTTCAATCTCAGAAATGCCAATCAG	487
QY	969	GTCAAGGTTTGCTACATTTGTATGTGTGATGCTTCTCCCAAGTATATTACTATA	1028
Db	488	GTCAAGGTTTGCTACATTTGTATGTGTGATGCTTCTCCCAAGTATATTACTATA	547
QY	1029	TTAAGAGTTTGCTGACAAACAGA	1051
Db	548	TTAAGAGTTTGCTGACAAACAGA	570
RESULT 13			
LOCUS	BX283642		
DEFINITION	BX283642 NIH_MGC_48 Homo sapiens cDNA clone IMAGE958K011707 ; IMAGE:4856568, mRNA sequence.		
ACCESSION	BX283642		
VERSION	BX283642.1		
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	Ebert, L., Hell, O., Hemig, S., Neubert, P., Parsch, E., Peters, M., Radelof, U., Schneider, D. and Korn, B.		
TITLE	Human Unigeneset - RZPD		
JOURNAL	Unpublished (2003)		
COMMENT	Contact: Ina Rolfs RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany Tel: +49 30 32638 101 Fax: +49 30 32639 111 www.rzpd.de		
FEATURES	This clone is available royalty-free from RZPD; contact RZPD (clone@rzpd.de) for further information. Seq primer: PCMV-M13v. Primer sequence: CCGTGTAAACGACGCGCCACT.		
SOURCE	Location/Qualifiers 1..499 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="IMAGE958K011707 ; IMAGE:4856568" /tissue_type="primary B-cells from tonsils (cell line)" /lab_host="DH10B (phage-resistant)" /clone_lib="NIH_MGC_48" /note="Organ: B-cells; Vector: pOT57; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Site-selected for average insert size 1.8kb. Library constructed by Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."		

Qy		72	GTCACACAGATGAATAATCTGTGCGAACAAGTGGATTAAAAAACAGTCCCTCAACTCT	133
Dd		62	GTCAACACCATGAAATATCTCTGCTGACACAGTGATAAAAAACAGTCCCTCAACTCT	121
Qy		132	TCTCTGTTTTATTTCTTCAACTCTCACTTCTTTAGAGTTTACAGAAAAATATTATATA	191
Dd		122	TCTCTGTTTTATTTCTTCAACTCTCACTTCTTTAGAGTTTACAGAAAAATATTATATA	181
Qy		192	CGACTCTTTAAAAGATCTATGTCTTGAAAATAGAGAAGAACACAGSTCTGGCCAAGGA	251
Dd		182	CGACTCTTTAAAAGATCTATGTCTTGAAAATAGAGAAGAACACAGSTCTGGCCAAGGA	241
Qy		252	CGTCTGCACCAATGTGTGTGAGTTTGAATGCAACATPFGCCCCCTACTGGGAATPACAGAACT	311
Dd		242	CGTCTGCACCAATGTGTGTGAGTTTGAATGCAACATGTGCCCTACTGGGAATPACAGAACT	301
Qy		312	GCAAGGACCTGGGAGCATTCTTAAGTGTCAACGTTTTTCTATGACTTTTAGTAGAGATGAG	371
Dd		302	GCAAGGACCTGGGAGCATTCTTAAGTGTCAACGTTTTTCTATGACTTTTAGTAGAGATGAG	361
Qy		372	AGCAGAGAGTAGATCTCTPAAAAAGCATGTGAGAGAGATCAAATGTTTTATATCAACATCC	431
Dd		362	AGCAGAGAGTAGATCTCTPAAAAAGCATGTGAGAGAGATCAAATGTTTTATATCAACATCC	421
Qy		432	TTTATTTATTGATTCATTTGATTAACAGTGTGTTTATGATATGATTTTCTATCTTTT	491
Dd		422	TTTATTTATTGATTCATTTGATTAACAGTGTGTTTATGATATGATTTTCTATCTTTT	481
Qy		492	CCCTTGACGTTTACTTTC 509	
Dd		482	CCCTTGACGTTTACTTTC 499	
RESULT_14				
Bg170756				
LOCUS			511 bp mRNA linear EST 06-FEB-2001	
DEFINITION			60332387.F1 NIH_MGC_89 Homo sapiens cDNA clone IMAGE:4426960 5'	
FEATURES			mRNA sequence.	
ACCESSION			Bg170756	
VERSION			Bg170756.1 GI:12677459	
KEYWORDS			EST.	
SOURCE			Homo sapiens (human)	
ORGANISM			Homo sapiens	
REFERENCE			Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS			Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
TITLE			1 (bases 1 to 511)	
JOURNAL			NIH-MGC http://mgc.nci.nih.gov/.	
COMMENT			National Institutes of Health, Mammalian Gene Collection (MGC)	
			Unpublished (1999)	
			Contact: Robert Strausberg, Ph.D.	
			Email: rgs@nih.nih.gov	
			Tissue Procurement: ATCC	
			cDNA Library Preparation: Life Technologies, Inc.	
			cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNLI)	
			DNA Sequencing by: Incyte Genomics, Inc.	
			Clone distribution: MGC clone distribution information can be	
			found through the I.M.A.G.E. Consortium/ILNL at:	
			http://image.lnl.gov	
			Plate: ILNM10174 row: n column: 17	
			High quality sequence stop: 511.	
FEATURES			Location/Qualifiers	
Source			1..511	

ORIGIN	Query Match	22.9%	Score 498	DB 13	Length 499
	Best Local Similarity	100.0%	Pred. No. 2.9e-63		
Matches	498	Conservative	0	Mismatches	0
				Indels	0
				Gaps	0
OY	12 GATTATGACCTTACGAGACGCACTTTCGTCCTTTGGGACCTTTGATAGCACTTCGAGGAT				71
Db	2 GATTATGACCTTACGAGACGCACTTTCGTCCTTTGGGACCTTTGATAGCACTTCGAGGAT				61

```

FEATURES
source
location/Qualifiers
1..511
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_id="IMAGE:4426960"
/tissue_type="hypertrophoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_id="NH MGCC 89"
/note="Organ: Kidney; Vector: pCMV-SPORT; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.3 kb. Library enriched for

```







GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 11, 2004, 17:20:29 ; Search time 66 seconds  
(without alignments)  
5019.783 Million cell updates/sec

Title: US-09-966-880A-7\_COPY\_80\_676

Sequence: 1 atgcagacagcctctgatgtaa.....ttcgtacttggagatttga 597

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: Issued Patents NA: \*  
1: /cgm2\_6/prodata/2/ina/5A COMB.seq: \*  
2: /cgm2\_6/prodata/2/ina/5B COMB.seq: \*  
3: /cgm2\_6/prodata/2/ina/5A COMB.seq: \*  
4: /cgm2\_6/prodata/2/ina/5B COMB.seq: \*  
5: /cgm2\_6/prodata/2/ina/PCTUS COMB.seq: \*  
6: /cgm2\_6/prodata/2/ina/backfile1.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	114.8	19.2	610	1	US-08-816-241-2 Sequence 2, Appl
2	114.8	19.2	610	1	US-09-128-395-2 Sequence 2, Appl
3	63.6	10.7	891	1	US-08-687-895-2 Sequence 2, Appl
4	63.6	10.7	891	2	US-09-040-482-2 Sequence 2, Appl
5	60	10.1	879	1	US-08-158-682A-3 Sequence 3, Appl
6	56.4	9.4	879	1	US-08-158-682A-1 Sequence 1, Appl
7	56.4	9.4	879	1	US-08-015-203-1 Sequence 1, Appl
8	38	6.4	1929	4	US-09-380-420C-1 Sequence 1, Appl
9	38	6.4	1929	4	US-09-899-642A-1 Sequence 1, Appl
10	37.8	6.3	922	3	US-09-189-060B-73 Sequence 73, Appl
11	36.6	6.1	4689	3	US-09-105-537-34 Sequence 34, Appl
12	36.6	6.1	36778	3	US-09-105-537-5 Sequence 5, Appl
13	36.6	6.1	38506	3	US-09-320-878-19 Sequence 19, Appl
14	36.6	6.1	38506	4	US-09-141-908-1 Sequence 1, Appl
15	36.6	6.1	38506	4	US-09-657-440-19 Sequence 19, Appl
16	36.4	6.1	624	4	US-09-252-991A-10053 Sequence 10053, A
17	36.4	6.1	1959	4	US-09-252-991A-9783 Sequence 9783, Ap
18	36.4	6.1	7218	1	US-08-232-463-14 Sequence 14, Appl
19	35.4	5.9	2190	4	US-09-015-188-1 Sequence 1, Appl
20	35.4	5.9	2584	4	US-09-716-129-47 Sequence 47, Appl
21	35	5.9	80246	3	US-09-078-294-4 Sequence 4, Appl
22	35	5.9	80595	3	US-09-078-294-3 Sequence 3, Appl
23	34.8	5.8	2505	4	US-09-334-818A-1 Sequence 117, Appl
24	34.6	5.8	1098	4	US-09-170-496D-117 Sequence 117, Appl
25	34.6	5.8	1098	4	US-09-170-496D-225 Sequence 225, Ap
26	34.6	5.8	1597	2	US-08-724-974A-1 Sequence 1, Appl
27	34.6	5.8	1697	4	US-09-364-425B-26 Sequence 26, Appl

28	34.6	5.8	2575	4	US-09-716-129-25	Sequence 25, Appl
29	34.4	5.8	2828	4	US-09-016-434-1458	Sequence 1458, Ap
30	34.4	5.8	2980	4	US-09-266-225D-11	Sequence 11, Appl
31	34.2	5.7	5159	2	US-08-146-930-3	Sequence 3, Appl
32	34.2	5.7	5159	3	US-08-458-240-3	Sequence 3, Appl
33	34.2	5.7	5159	5	PCT-US93-03993-3	Sequence 3, Appl
34	34.2	5.7	6359	4	US-09-475-252-1	Sequence 1, Appl
35	33.8	5.7	417	4	US-09-252-991A-9297	Sequence 9297, Ap
36	33.8	5.7	1215	4	US-09-252-991A-9259	Sequence 9259, Ap
37	33.8	5.7	1488	4	US-09-252-991A-9307	Sequence 9307, Ap
38	33.8	5.7	1602	4	US-09-252-991A-9270	Sequence 9270, Ap
39	33.8	5.7	15872	3	US-08-105-537-1	Sequence 1, Appl
40	33.8	5.7	15872	4	US-09-091-609-1	Sequence 1, Appl
41	33.8	5.7	15872	4	US-09-091-609-3	Sequence 3, Appl
42	33.8	5.7	4403765	3	US-09-103-840A-2	Sequence 2, Appl
43	33.8	5.7	4411529	3	US-09-103-840A-1	Sequence 1, Appl
44	33.6	5.6	1683	4	US-09-252-991A-8337	Sequence 8337, Ap
45	33.6	5.6	1833	4	US-09-252-991A-8371	Sequence 8371, Ap

## ALIGNMENTS

RESULT 1  
US-08-816-241-2  
Sequence 2, Application US/08816241  
Patent No. 5804185  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Goll, Surya K.  
TITLE OF INVENTION: NOVEL RNA EDITING ENZYME  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/816,241  
FILING DATE: Filed Herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-845-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 610 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: PROSTUT09  
CLONE: 1646833  
US-08-816-241-2  
Query Match 19.2%; Score 114.8; DB 1; Length 610;  
Best Local Similarity 53.9%; Pred. No. 1,76-25;  
Matches 289; Conservative 0; Mismatches 223; Indels 15; Gaps 2;



QY 16 ATGACCGGAGGAGGAGTTCTTTACCAATCAAAAATGTCGGCTGAGGCTGCGGCT 75  
DB 34 ATGTATCCAGGACATTTCTACTTCAATTTAAACCTATGGAAAGCCACATCGGAC 93  
QY 76 GAGACCTACCTGTGCTACGATGAGAGGCGGTGACATGCTACATCTTTTCACTGAC 135  
DB 94 GAAACTTGCTGTGCTTCAACCGTGAAGATATTAAGCCGCGCTCATGTTCTCTGGAAG 153  
QY 136 TTGTTATCTTTCGCA-----TAAAGACGGCTGCCAGTGAATGCTCTTC 183  
DB 154 ACGGCGGTCTTCCGAACAGGTGATCTGTGAGACCATTTGATGCAAGAGGCTGCTTC 213  
QY 184 CTCGCGTACATCTCGACCTGAGACCTAGACCCCTGCGCGCTGCTACCGCGCTGCTTC 243  
DB 214 CTCTCTGTTCTGCGACGACATATCTGCTCTTAACACAAAGTACAGGTCTGCTGTC 273  
QY 244 ACCTCTGAGCCCTCTGCTACGATGCTGCGCGCATGTGCGCGACTTTCTGCGAGGAC 303  
DB 274 ACATCTTGAGCCCTTGCACAGCTGTGAGGAGGTGCGGAGTCTCTGCGCGACGAC 333  
QY 304 CCCACCTCAGTCTGAGATCTTACCGCGCGCTCTACTTGTGAGACCGCAAGCT 363  
DB 334 AGCAAGTGAATCTCACTCTTCAACCGCGCTCTACTCTCAAGTATC---CATGT 390  
QY 364 GAGCCGAGGGGCTGCGGCGCTGACCGCGCGGGGTGCAATAGCCATCATGACCTTC 423  
DB 391 TACCAAGAGGGGCTCGCAGCCCTGAGTCAAGAGGGGTGCTGTGAGTCAATGACAT 450  
QY 424 AAGATTTATTTTACTGCTGAATTAATTTGTAGAAACATGAAAGAACTTTCAAGCC 483  
DB 451 GAAATTTTAAATATTTGTGGGAAACTTTGTGTACATATGATAGCCATTCAGCTT 510  
QY 484 TGGAGAGGGCTGCATGAAATTCAGTTCGTCTCCAGACAGCTTCGCGCATCTCT 539  
DB 511 TGGAGGATTTAAACCACTTTCGACTTGTGAAAGAGGCTACGGAGAGTCT 566

## RESULT 2

US-09-128-395-2  
Sequence 2, Application US/09128395  
Patent No. 6087108  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Goli, Surya K.  
TITLE OF INVENTION: NOVEL RNA EDITING ENZYME  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/128,395  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/815,241  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: BP-0239 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:  
LENGTH: 610 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: PROSTUT09  
CLONE: 1646833

## US-09-128-395-2

Query Match 19.2%; Score 114.8; DB 3; Length 610;  
Best Local Similarity 53.9%; Pred. No. 1.7e-25;  
Matches 289; Conservative 0; Mismatches 232; Indels 15; Gaps 2;

QY 16 ATGACCGGAGGAGGAGTTCTTTACCAATCAAAAATGTCGGCTGAGGCTGCGGCT 75  
DB 34 ATGTATCCAGGACATTTCTACTTCAATTTAAACCTATGGAAAGCCACATCGGAC 93  
QY 76 GAGACCTACCTGTGCTACGATGAGAGGCGGTGACATGCTACATCTTTTCACTGAC 135  
DB 94 GAAACTTGCTGTGCTTCAACCGTGAAGATATTAAGCCGCGCTCATGTTCTCTGGAAG 153  
QY 136 TTGTTATCTTTCGCA-----TAAAGACGGCTGCCAGTGAATGCTCTTC 183  
DB 154 ACGGCGGTCTTCCGAACAGGTGATCTGTGAGACCATTTGATGCAAGAGGCTGCTTC 213  
QY 184 CTCGCGTACATCTCGACCTGAGACCTAGACCCCTGCGCGCTGCTACCGCGCTGCTTC 243  
DB 214 CTCTCTGTTCTGCGACGACATATCTGCTCTTAACACAAAGTACAGGTCTGCTGTC 273  
QY 244 ACCTCTGAGCCCTCTGCTACGATGCTGCGCGCATGTGCGCGACTTTCTGCGAGGAC 303  
DB 274 ACATCTTGAGCCCTTGCACAGCTGTGAGGAGGTGCGGAGTCTCTGCGCGACGAC 333  
QY 304 CCCACCTCAGTCTGAGATCTTCAACCGCGCGCTCTACTTGTGAGACCGCAAGCT 363  
DB 334 AGCAAGTGAATCTCACTCTTCAACCGCGCTCTACTCTCAAGTATC---CATGT 390  
QY 364 GAGCCGAGGGGCTGCGGCGCTGACCGCGCGGGGTGCAATAGCCATCATGACCTTC 423  
DB 391 TACCAAGAGGGGCTCGCAGCCCTGAGTCAAGAGGGGTGCTGTGAGTCAATGACAT 450  
QY 424 AAGATTTATTTTACTGCTGAATTAATTTGTAGAAACATGAAAGAACTTTCAAGCC 483  
DB 451 GAAATTTTAAATATTTGTGGGAAACTTTGTGTACATATGATAGCCATTCAGCTT 510  
QY 484 TGGAGAGGGCTGCATGAAATTCAGTTCGTCTCCAGACAGCTTCGCGCATCTCT 539  
DB 511 TGGAGGATTTAAACCACTTTCGACTTGTGAAAGAGGCTACGGAGAGTCT 566

## RESULT 3

US-08-687-895-2  
Sequence 2, Application US/08687895  
Patent No. 5747319  
GENERAL INFORMATION:  
APPLICANT: Au-Young, Janice  
APPLICANT: Hawkins, Phillip R.  
TITLE OF INVENTION: A NOVEL HUMAN RNA EDITING ENZYME  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: U.S.  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5



CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/687,895  
 FILING DATE: Filed Herewith  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Billings, Lucy J.  
 REGISTRATION NUMBER: 36,749  
 REFERENCE/DOCKET NUMBER: PF-0109 US  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415-855-0555  
 TELEFAX: 415-845-4166  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 891 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 IMMEDIATE SOURCE:  
 LIBRARY: MUSCNOT1  
 CLONE: 57953  
 US-08-687-895-2

Query Match 10.7%; Score 63.6; DB 1; Length 891;  
 Best Local Similarity 47.0%; Pred. No. 1.1e-09;  
 Matches 223; Conservative 2; Mismatches 246; Indels 3; Gaps 1;

QY 31 TTCTTTACCAATTCAAAAATGTCGGCTGAGGCTGAGGCTGAGCTTACTGTC 90  
 DB 306 TTAAATTCAGATCCCGAATGTGAGATCAATTCGGGAGGAACAAGACCTTCTGTC 365  
 QY 91 TAGCTAGTGAAGAGCGGTGACAGTGTACATCTTTTCACTGAGACTTGGTATCTTCC 150  
 DB 366 TATGTGTTGAAGACACAGGCGCAAGGGGCGCAAGTGCAGGCAATCTCGGGATACCTTAG 425  
 QY 151 AATAAGAACGGCTGCGACAGTGAATTGCTCTTCTCCGCTACATCTCGAGCTGGACCTA 210  
 DB 426 GATGAGCATGCGGTGCTCCATGACAGAGAGCTTTCTTCAACACATCTCCGACGCTTC 485  
 QY 211 GACCTGAGCGGTGCTACCGGCTGACCTGCTGACCTCTGAGCCCTGCTACGACTGT 270  
 DB 486 GACCAAGCCTGCGGTACATGTCACCTGTATGTGTCTCCAGCCCTGTGACGGTGT 545  
 QY 271 GCCCGAATGTGGCCGACTTTCGAGAGGAAACCCCACTCATGTCTGAGATCTTCACC 330  
 DB 546 GCTGACCGCATTTACAAACCTTAGCAACCAAGAACCTGTGCTCATTTCTGTG 605  
 QY 331 GCGGCGCTTACTTCTGTGAGGACCGGCAAGGCTGAGCCGAGGGGCTGCGGGCTGCAC 390  
 DB 606 GGTGACTCTTCAATGTGGAGAGCCGAGATCCAG--GCTGCTTGAAAGACTGAAG 662  
 QY 391 CGCGCCGGGTGCAATAGCCATCATGACCTTCAAGATTAATTTTATCTGCTGAATACT 450  
 DB 663 GAGGCTGCTGTAACTGCGCATCATGAGCCCAAGACTTGAATATGTCTGGCAGAT 722  
 QY 451 TTGTGAAAAACCATGAAGAAGACTTTCAAGCTGAGAGGGCTGCATGAAT 504  
 DB 723 TTGTGAGAGCAAGAGGGTGAATCAAGGCTTTCAACCTTGGAGGACATT 776

RESULT 4  
 US-09-040-482-2  
 Sequence 2, Application US/09040482

Patent No. 591556  
 GENERAL INFORMATION:  
 APPLICANT: Au-Young, Janice  
 APPLICANT: Hawkins, Phillip R.  
 APPLICANT: Hillman, Jennifer L.  
 TITLE OF INVENTION: A NOVEL HUMAN MRNA EDITING ENZYME  
 NUMBER OF SEQUENCES: 5  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Inocyte Pharmaceuticals, Inc.  
 STREET: 3174 Porter Drive  
 CITY: Palo Alto

STATE: CA  
 COUNTRY: U.S.  
 ZIP: 94304  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FastSeq Version 1.5  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/040,482  
 FILING DATE:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/687,895  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Billings, Lucy J.  
 REGISTRATION NUMBER: 36,749  
 REFERENCE/DOCKET NUMBER: PF-0109 US  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 415-855-0555  
 TELEFAX: 415-845-4166  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 891 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 IMMEDIATE SOURCE:  
 LIBRARY: MUSCNOT1  
 CLONE: 57953  
 US-09-040-482-2

Query Match 10.7%; Score 63.6; DB 2; Length 891;  
 Best Local Similarity 47.0%; Pred. No. 1.1e-09;  
 Matches 223; Conservative 2; Mismatches 246; Indels 3; Gaps 1;

QY 31 TTCTTTACCAATTCAAAAATGTCGGCTGAGGCTGAGGCTGAGCTTACTGTC 90  
 DB 306 TTAAATTCAGATCCCGAATGTGAGATCAATTCGGGAGGAACAAGACCTTCTGTC 365  
 QY 91 TAGCTAGTGAAGAGCGGTGACAGTGTACATCTTTTCACTGAGACTTGGTATCTTCC 150  
 DB 366 TATGTGTTGAAGACACAGGCGCAAGGGGCGCAAGTGCAGGCAATCTCGGGATACCTTAG 425  
 QY 151 AATAAGAACGGCTGCGACAGTGAATTGCTCTTCTCCGCTACATCTCGAGCTGGACCTA 210  
 DB 426 GATGAGCATGCGGTGCTCCATGACAGAGAGCTTTCTTCAACACATCTCCGACGCTTC 485  
 QY 211 GACCTGAGCGGTGCTACCGGCTGACCTGCTGACCTCTGAGCCCTGCTACGACTGT 270  
 DB 486 GACCAAGCCTGCGGTACATGTCACCTGTATGTGTCTCCAGCCCTGTGACGGTGT 545  
 QY 271 GCCCGAATGTGGCCGACTTTCGAGAGGAAACCCCACTCATGTCTGAGATCTTCACC 330  
 DB 546 GCTGACCGCATTTACAAACCTTAGCAACCAAGAACCTGTGCTCATTTCTGTG 605  
 QY 331 GCGGCGCTTACTTCTGTGAGGACCGGCAAGGCTGAGCCGAGGGGCTGCGGGCTGCAC 390  
 DB 606 GGTGACTCTTCAATGTGGAGAGCCGAGATCCAG--GCTGCTTGAAAGACTGAAG 662  
 QY 391 CGCGCCGGGTGCAATAGCCATCATGACCTTCAAGATTAATTTTATCTGCTGAATACT 450  
 DB 663 GAGGCTGCTGTAACTGCGCATCATGAGCCCAAGACTTGAATATGTCTGGCAGAT 722  
 QY 451 TTGTGAAAAACCATGAAGAAGACTTTCAAGCTGAGAGGGCTGCATGAAT 504  
 DB 723 TTGTGAGAGCAAGAGGGTGAATCAAGGCTTTCAACCTTGGAGGACATT 776

RESULT 5  
 US-08-158-682A-3  
 Sequence 3, Application US/08158682A



```
Patent No. 5434058
GENERAL INFORMATION:
APPLICANT: Davidson, Nicholas O.
TITLE OF INVENTION: Apolipoprotein B RNA Editing Protein:
NUMBER OF SEQUENCES: 18
TITLE OF INVENTION: Composition and Method
CORRESPONDENCE ADDRESS:
ADDRESS: ARNOLD, WHITE & DURKEE
STREET: 321 No. 5434058th Clark Street, Suite 800
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60610
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/158,682A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Cooley, Ronald B.
REGISTRATION NUMBER: 27,187
REFERENCE/DOCKET NUMBER: ARCD:085
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 744-0090
TELEFAX: (312) 245-4961
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 879 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 23...730
US-08-158-682A-3

Query Match          10.1%; Score 60; DB 1; Length 879;
Best Local Similarity 54.1%; Pred. No. 1.4e-08;
Matches 145; Conservative 0; Mismatches 120; Indels 3; Gaps 1;
```

```

TITLE OF INVENTION: Composition and Method
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESS: ARNOLD, WHITE & DURKEE
STREET: 321 No. 5434058th Clark Street, Suite 800
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60610
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/158,682A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Cooley, Ronald B.
REGISTRATION NUMBER: 27,187
REFERENCE/DOCKET NUMBER: ARCD:085
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 744-0090
TELEFAX: (312) 245-4961
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 879 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-158-682A-1

Query Match          9.4%; Score 56.4; DB 1; Length 879;
Best Local Similarity 53.4%; Pred. No. 1.7e-07;
Matches 142; Conservative 0; Mismatches 121; Indels 3; Gaps 1;
```



CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60610  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/015,203  
FILING DATE: 19930209  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Coolley, Ronald B.  
REGISTRATION NUMBER: 27,187  
REFERENCE/DOCKET NUMBER: ARCD:069  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 744-0090  
TELEFAX: (312) 245-4961  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 879 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-015-203-1

Query Match 9.4%; Score 56.4; DB 1; Length 879;  
Best Local Similarity 53.4%; Pred. No. 1.7e-07;  
Matches 142; Conservative 0; Mismatches 121; Indels 3; Gaps 1;  
QY 197 CGAGTGGAGCTAGACCTGCGCGCTGCTACCGGCTACCTGCTTCACTCTCGAGACC 256  
DB 255 CAGAAAGATACCTTTGTCCAAACACAGATGCTCCATTAACGTCTGCTGCTGATC 314  
QY 257 CCTGCTACGACTGTGCGGACATGTGGCCGACTTTCTGCGAGGAGACCCCAACTCAGTC 316  
DB 315 CCGTGGGAGTGGCTCAGAGGCGCATTAACGAAATTTTGAAGCCGATACCCCATGTAATCTC 374  
QY 317 TGAGGATCTTCCCGGCGCTCTACTCTGTGAGAGACCGAAGCTGAGCCCGAGGCGC 376  
DB 375 TGTATTATTAATAGACGCGCTTTATCAACGACGATCTCGAATGCGC---AAGGAC 431  
QY 377 TCGCGCGCTGACCGCGCGCGGAGTGAATAGCATCATGATCCTTCAAGATTATTTT 436  
DB 432 TCAGGCACTTATTTAGCAGCGGTGTACTATCAATCATGACGAGCAAGAGTCTGCT 491  
QY 437 ACTGCTGGAATCTTTGTAGAAAC 462  
DB 492 ACTGCTGAGGAATTTGTCACTAC 517

RESULT 8  
US-09-380-420C-1  
Sequence 1, Application US/09380420C  
Patent No. 6300544  
GENERAL INFORMATION:  
APPLICANT: Halkier, Barbara  
Kahn, Rachel  
Bak, Soren  
Moller, Birger  
TITLE OF INVENTION: Cytochrome P450 Monooxygenases  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Syngenta Patent Dept.  
STREET: 3054 Cornwallis Road  
CITY: RTP  
STATE: NC  
COUNTRY: USA  
ZIP: 27709  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/380,420C  
FILING DATE: 12-No. 6300544-1999  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Weigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: S-21251A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8587  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1929 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
IMMEDIATE SOURCE:  
CLONE: P450x  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 81..1673  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-380-420C-1

Query Match 6.4%; Score 38; DB 4; Length 1929;  
Best Local Similarity 53.3%; Pred. No. 0.11; 70; Indels 0; Gaps 0;  
Matches 80; Conservative 0; Mismatches 70; Indels 0;  
QY 236 CTGCTTCACTCTCTGAGCGCTGCTACGACTGTGCGGACATGTGGCGGACTTCTGC 295  
DB 793 CAGCTTCTCGCGGAGACTTTTCCCAACGCGCGCGCTCGCGGACGCTCT 852  
QY 296 GAGGGAACCCCACTCAGTCTGAGATCTTCAACCGCGCGCTTCACTTGTGAGACC 355  
DB 853 CGGGCTTCTCGCGCGCGCGGAGCGCATCTTCAACGAGCTCGACGCTTCTTGAAGAAG 912  
QY 356 GCAAGCTGAGCGCGGCGCTCGCGCGC 385  
DB 913 TCATGACGACGACATGACCGCGCGCGC 942

RESULT 9  
US-09-899-642A-1  
Sequence 1, Application US/09899642A  
Patent No. 6649814  
GENERAL INFORMATION:  
APPLICANT: Halkier, Barbara  
Kahn, Rachel  
Bak, Soren  
Moller, Birger  
TITLE OF INVENTION: Cytochrome P450 Monooxygenases  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Syngenta Patent Dept.  
STREET: 3054 Cornwallis Road  
CITY: RTP  
STATE: NC  
COUNTRY: USA  
ZIP: 27709  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/899,642A  
FILING DATE: 05-Jul-2001  
CLASSIFICATION: <Unknown>



PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/380,420  
FILING DATE: 12-NO. 6649814-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: S-21251A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8587  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1929 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
IMMEDIATE SOURCE:  
CLONE: P4500x  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 81..1673  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-899-642A-1

Query Match  
Best Local Similarity 6.4%; Score 38; DB 4; Length 1929;  
Best Local Similarity 53.3%; Pred. No. 0.11;  
Matches 80; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 236 CCGGCTTCAACCTCCGAGAGCCCTGTACAGTGGCCGAGCATGGCCGCACTTTTCGC 295  
DB 793 CCAAGCTTCTCCGCGAGACTTCTTCCCAAGCCGCGCCGCTCGCCGACCGCTCT 852  
QY 296 GAGGGAACCCCAACTCACTGAGATCTTCAACCGCGCCCTCTACTTGTGTGAGACC 355  
DB 853 CGGGCTTCTCGCCCGCCGCGAGCGCATTTCAAGAGCTGACGCTTCTTCCGAGAGG 912

QY 356 GCAAGGCTGAGCCCGAGGGGCTGGCGGCGC 385  
DB 913 TCATGACCAAGCATGAGACCGCGCGCGCC 942

RESULT 10  
US-09-189-060B-73  
Sequence 73, Application US/09189060B  
Patent No. 6270968  
GENERAL INFORMATION:  
APPLICANT: Dalboge, Henrik  
APPLICANT: Sandel, Thomas  
APPLICANT: Kauppinen, Markus  
APPLICANT: Borge, Diderichsen  
TITLE OF INVENTION: Method Of Providing No. 6270968e1 DNA Sequences  
FILE REFERENCE: 4772.204-US  
CURRENT APPLICATION NUMBER: US/09/189,060B  
CURRENT FILING DATE: 1998-11-10  
PRIOR APPLICATION NUMBER: PCT/DK97/00216  
PRIOR FILING DATE: 1997-05-12  
NUMBER OF SEQ ID NOS: 74  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 73  
LENGTH: 922  
TYPE: DNA  
ORGANISM: Hybrid  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (2)...(922)  
US-09-189-060B-73

Query Match  
Best Local Similarity 6.3%; Score 37.8; DB 3; Length 922;  
Best Local Similarity 49.3%; Pred. No. 0.091;  
Matches 99; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 176 TGCTCTTCTCCCTCAATCTCGAGTGGAGCTAGACCTTGCGCGCTGCTACCGGCTCA 235

DB 378 TGATGCTCCAGTCCACCAACAGGGCGGAGCTGTCCGACCAACCACTTGCTCA 437  
QY 236 CCGGTTTCACTTCTGAGAGCCCTGTGACAGTGTGCGGAGATGAGCCGCACTTCTGC 295  
DB 438 TCCCCCGCGCGCGCTGCGGATTTGACGAGCTGACCTTCCAGTACGAGCCGCTTCG 497  
QY 296 GAGGGAACCCCAACTCACTGAGATCTTCAACCGCGCCCTCTACTTGTGTGAGACC 355  
DB 498 CCGCGCCGAGTACGCGCGGCGGTGAGCTCCCGCGCGAGTGCACAGATGCCGAGGCCA 557

QY 356 GCAAGGCTGAGCCCGAGGCGC 376  
DB 558 TCAAGCGGCGCTCCAGTGC 578

RESULT 11  
US-09-105-537-34  
Sequence 34, Application US/09105537A  
Patent No. 6265202  
GENERAL INFORMATION:  
APPLICANT: Sherman, D.H.  
APPLICANT: Liu, H.  
APPLICANT: Xue, Y.  
APPLICANT: Zhao, L.  
TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
FILE REFERENCE: 600.438US1  
CURRENT APPLICATION NUMBER: US/09/105,537A  
CURRENT FILING DATE: 1998-06-26  
NUMBER OF SEQ ID NOS: 43  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 34  
LENGTH: 4689  
TYPE: DNA  
ORGANISM: Streptomyces venezuelae  
US-09-105-537-34

Query Match  
Best Local Similarity 6.1%; Score 36.6; DB 3; Length 4689;  
Best Local Similarity 46.9%; Pred. No. 0.48;  
Matches 114; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 156 GAAGCGCTGACGAGTGAATTGCTTCTCCGCTACATCTGAGACTGAGACCTAGACC 215  
DB 3402 GATGCCCCACACGAGAGCCGAAACACTCTCTCTGTGACGCGAGGGGGAACACCCC 3461  
QY 216 TGCCCGTGTCTACCGCTCACTGATTCACCTTCGAGGCCCTGTACGATGTGCCG 275  
DB 3462 CGAGGCCACCACTACCGCGCACTCAACCGCATCGGGGCGCGGTCAACCATCGCCG 3521  
QY 276 ACATGTGCGGACTTTTGTGAGAGGAACCCCACTTCACTGATGTAGATCTTACCGCGG 335  
DB 3522 CTGGGAGCTGCGGACCCCGACGCGCATGCGGACCTTCTGAGCGCATCCCGCGAGAC 3581  
QY 336 CTTCTACTTGTGTGAGGACCGGAGGCTGAGCCCGGAGGCTCGGCGGCTGACCGCG 395  
DB 3582 GCCCTTACCGCGGTCTTCAACCGCGCGGACCGGCGGAGATCGCTGACGTCAAC 3641

QY 396 CGG 398  
DB 3642 CGG 3644

RESULT 12  
US-09-105-537-5  
Sequence 5, Application US/09105537A  
Patent No. 6265202  
GENERAL INFORMATION:  
APPLICANT: Sherman, D.H.  
APPLICANT: Liu, H.  
APPLICANT: Xue, Y.  
APPLICANT: Zhao, L.  
TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
FILE REFERENCE: 600.438US1  
CURRENT APPLICATION NUMBER: US/09/105,537A



; CURRENT FILING DATE: 1998-06-26  
 ; NUMBER OF SEQ ID NOS: 43  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 5  
 ; LENGTH: 36778  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces venezuelae  
 ; US-09-105-537-5

Query Match 6.1%; Score 36.6; DB 3; Length 36778;  
 Best Local Similarity 46.9%; Pred. No. 1.4;  
 Matches 114; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 156 GAACGGCTGCGACGTGGAATGCTCTTCCTCCGCTACATCTCGAGCTGGAGCTAGACCC 215  
 DB 30392 GATGGCCACCAACCGAGCCGAAACCTCTCTCTCTGTCAGCCGAGCGGCAAGACCC 30451  
 QY 216 TGGCCGCTGTACCGCGTCACTGTTCACCTCTGTAGAGCCCTGTACGACTGTGCGG 275  
 DB 30452 CGGAGCCACCACTCAACCGCCGACGATCGGCGCGCTACACCTGCGCG 30511  
 QY 276 ACATGTGCGCACTTCTGTGAGAGAAACCCCACTTCAGTGAAGATCTTACCGCGCG 335  
 DB 30512 CTGGAGCGTGGCGACCCCGACGCGCATGCGCTCTGACGCGCATCCCGCGAGAC 30571  
 QY 336 CCTTACTTCTGTGAGAGACCGCAAGCTGAGCCCGAGGGCTGCGCGCTGACCGCGC 395  
 DB 30572 GCGCCCTACCGCGCTGCTCCACACCGCGGCGGACCGGGCGGAGTCCGTGACGTCAC 30631  
 QY 396 CGG 398  
 DB 30632 CGG 30634

## RESULT 13

US-09-320-878-19  
 ; Sequence 19; Application US/09320878A  
 ; Patent No. 6117659

; GENERAL INFORMATION:  
 ; APPLICANT: ASHLEY, Gary  
 ; APPLICANT: BETLACH, Melanie C.  
 ; APPLICANT: BETLACH, Mary C.  
 ; APPLICANT: MCDANIEL, Robert  
 ; APPLICANT: TANG, Li  
 ; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE  
 ; FILE REFERENCE: 30622002120  
 ; CURRENT APPLICATION NUMBER: US/09/320,878A  
 ; EARLIER FILING DATE: 1998-05-27  
 ; EARLIER APPLICATION NUMBER: CIP OF 09/141,908  
 ; EARLIER FILING DATE: 1998-08-28  
 ; EARLIER APPLICATION NUMBER: CIP OF 09/073,538  
 ; EARLIER FILING DATE: 1998-05-06  
 ; EARLIER APPLICATION NUMBER: CIP OF 08/846,247  
 ; EARLIER FILING DATE: 1997-04-30  
 ; EARLIER APPLICATION NUMBER: 60/119,139  
 ; EARLIER FILING DATE: 1999-02-08  
 ; EARLIER APPLICATION NUMBER: 60/100,880  
 ; EARLIER FILING DATE: 1998-09-22  
 ; EARLIER APPLICATION NUMBER: 60/087,080  
 ; EARLIER FILING DATE: 1998-05-28  
 ; NUMBER OF SEQ ID NOS: 34  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 19  
 ; LENGTH: 38506  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces venezuelae  
 ; US-09-320-878-19

Query Match 6.1%; Score 36.6; DB 3; Length 38506;  
 Best Local Similarity 46.9%; Pred. No. 1.4; Indels 129; Indels 0; Gaps 0;  
 Matches 114; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 156 GAACGGCTGCGACGTGGAATGCTCTTCCTCCGCTACATCTCGAGCTGGAGCTAGACCC 215

DB 28534 GATGGCCACCAACCGAGCCGAAACCTCTCTCTCTGTCAGCCGAGCGGCAAGACCC 28593  
 QY 216 TGGCCGCTGTACCGCGTCACTGTTCACCTCTGTAGAGCCCTGTGACGACTGTGCGG 275  
 DB 28594 CGGAGCCACCACTCAACCGCGGAACTCAACCGCATCGGCGCGCGCTGACCATCGCGC 28653  
 QY 276 ACATGTGCGCACTTCTGTGAGAGAAACCCCACTTCAGTGAAGATCTTACCGCGCG 335  
 DB 28654 CTGGAGCGTGGCGACCCCGACGCGCATGCGACCTCTGTGACGCGCATCCCGCGAGAC 28713  
 QY 336 CCTTACTTCTGTGAGAGACCGCAAGCTGAGCCCGAGGGCTGCGCGCTGACCGCGC 395  
 DB 28714 GCGCCCTACCGCGCTGCTCCACACCGCGGCGGACCGGGCGGAGTCCGTGACGTCAC 28773  
 QY 396 CGG 398  
 DB 28774 CGG 28776

## RESULT 14

US-09-141-908-1  
 ; Sequence 1; Application US/09141908  
 ; Patent No. 6503741

; GENERAL INFORMATION:  
 ; APPLICANT: ASHLEY, Gary  
 ; APPLICANT: BETLACH, Melanie C.  
 ; APPLICANT: BETLACH, Mary  
 ; APPLICANT: MCDANIEL, Robert  
 ; APPLICANT: TANG, Li  
 ; TITLE OF INVENTION: Combinatorial Polyketide Libraries Produced Using a  
 ; FILE REFERENCE: Modular PKS Gene Cluster as Scaffold  
 ; CURRENT APPLICATION NUMBER: US/09/141,908  
 ; EARLIER FILING DATE: 1998-08-28  
 ; EARLIER APPLICATION NUMBER: CIP OF 09/073,538  
 ; EARLIER FILING DATE: 1998-05-06  
 ; EARLIER APPLICATION NUMBER: CIP OF 08/846,247  
 ; EARLIER FILING DATE: 1997-04-30  
 ; EARLIER APPLICATION NUMBER: PROV. 60/076,919  
 ; EARLIER FILING DATE: 1998-03-05  
 ; EARLIER APPLICATION NUMBER: PROV. 60/087,080  
 ; EARLIER FILING DATE: 1998-05-28  
 ; NUMBER OF SEQ ID NOS: 31  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 1  
 ; LENGTH: 38506  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces venezuelae  
 ; US-09-141-908-1

Query Match 6.1%; Score 36.6; DB 4; Length 38506;  
 Best Local Similarity 46.9%; Pred. No. 1.4;  
 Matches 114; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 156 GAACGGCTGCGACGTGGAATGCTCTTCCTCCGCTACATCTCGAGCTGGAGCTAGACCC 215  
 DB 28534 GATGGCCACCAACCGAGCCGAAACCTCTCTCTGTCAGCCGAGCGGCAAGACCC 28593  
 QY 216 TGGCCGCTGTACCGCGTCACTGTTCACCTCTGTAGAGCCCTGTGACGACTGTGCGG 275  
 DB 28594 CGGAGCCACCACTCAACCGCGGAACTCAACCGCATCGGCGCGCGCTGACCATCGCGC 28653  
 QY 276 ACATGTGCGCACTTCTGTGAGAGAAACCCCACTTCAGTGAAGATCTTACCGCGCG 335  
 DB 28654 CTGGAGCGTGGCGACCCCGACGCGCATGCGACCTCTGTGACGCGCATCCCGCGAGAC 28713  
 QY 336 CCTTACTTCTGTGAGAGACCGCAAGCTGAGCCCGAGGGCTGCGCGCTGACCGCGC 395  
 DB 28714 GCGCCCTACCGCGCTGCTCCACACCGCGGCGGACCGGGCGGAGTCCGTGACGTCAC 28773  
 QY 396 CGG 398



Db 28774 CGG 28776

RESULT 15  
US-09-657-440-19  
; Sequence 19; Application US/09657440  
; Patent No. 6509455  
; GENERAL INFORMATION:  
; APPLICANT: ASHLEY, Gary  
; APPLICANT: BETLACH, Melanie C.  
; APPLICANT: BETLACH, Mary C.  
; APPLICANT: MCDANIEL, Robert  
; APPLICANT: TANG, Li  
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE  
; FILE REFERENCE: 300622002120  
; CURRENT APPLICATION NUMBER: US/09/657,440  
; CURRENT FILING DATE: 2000-09-07  
; PRIOR APPLICATION NUMBER: 09/320,878  
; PRIOR FILING DATE: 1999-05-27  
; PRIOR APPLICATION NUMBER: CIP OF 09/141,908  
; PRIOR FILING DATE: 1998-08-28  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 19  
; LENGTH: 38506  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-09-657-440-19

Query Match 6.1%; Score 36.6; DB 4; Length 38506;  
Best Local Similarity 46.9%; Pred. No. 1.4;  
Matches 114; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY	156	GAACGGCTGCACGCTGGAATTGCTCTCTCCCGCTACATCGAGCTGGAGCTAGACCC	215
Db	28534	GATGGCCACCCAGGAGCCGAAACCTCTCTCTGTCAGCGACGCGGAAACAGCCCC	28593
QY	216	TGCGCGCTGCTACCGCGCTCACTGATTCACCTCCTGGAGCCCTGCTACGACTGTGCCG	275
Db	28594	CGGAGCCACCACTCACCGCCGAACTACCGCATGGGCGCCCGCTACCATGCCCC	28653
QY	276	ACATGTGCGGACTTCTGCGAGGAAACCCCAACTCAGTGTAGATCTTCAACCGCGC	335
Db	28654	CTGCGAGTGGCCGACCCGACCGCATGCGCACCTCTCGACGCACTCCCGCGAGAC	28713
QY	336	CCTTACTTCTGTGAGAGCGCAAGGCTGAGCCCGAGGGGCTGGGCGGCTGCACCGCG	395
Db	28714	GCCCCCTACCGCCGTCGTCCACACCGCGGCGACCGGGCGGCGATCCCTGAGCTCAC	28773
QY	396	CGG 398	
Db	28774	CGG 28776	

Search completed: March 12, 2004, 09:40:12  
Job time : 74 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 12, 2004, 09:15:07 ; Search time 248 Seconds  
(without alignments)  
8862.556 Million cell updates/sec

Title: US-09-966-880A-7\_COPY\_80\_676

Perfect score: 597  
Sequence: 1 atggacacgactcttgatgaa.....ttcgracttggacttga 597

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2432557 seqs, 1840798864 residues

Total number of hits satisfying Chosen parameters: 4865114

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications NA.\*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
17: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
18: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	597	100.0	2818	9	US-09-966-880A-7
2	475.4	79.6	2440	9	US-09-966-880A-1
3	273.4	45.8	6564	9	US-09-966-880A-10
4	273.4	45.8	11204	9	US-09-966-880A-35
5	271	45.4	271	9	US-09-966-880A-13
6	148	24.8	148	9	US-09-966-880A-12
7	118.4	19.8	1155	15	US-10-460-923-1
8	118.4	19.8	1534	9	US-09-729-674-173
9	114.8	19.2	987	9	US-09-925-300-999
10	114.8	19.2	1143	13	US-10-098-841-268
11	114.4	19.2	116	9	US-09-966-880A-14
12	110.4	18.5	1348	9	US-09-880-107-3282
13	107.2	18.0	823	15	US-10-247-671-62
14	107.2	18.0	1008	15	US-10-131-827-8890
15	104.4	17.5	2151	9	US-09-822-830A-359

16	90.2	15.1	2913	15	US-10-027-632-112023
17	90.2	15.1	2913	15	US-10-027-632-112024
18	90.2	15.1	201143	12	US-10-240-925-1039
19	78.4	13.1	371	10	US-09-918-995-30237
20	71.6	12.0	1404	14	US-10-198-846-13472
21	71.6	12.0	2773	15	US-10-104-047-1759
22	70.6	11.8	456	14	US-10-029-386-25187
23	70.2	11.8	390	9	US-09-786-692-9051
24	70.2	11.8	390	14	US-10-040-862-9051
25	70.2	11.8	390	15	US-10-057-475B-9051
26	70.2	11.8	390	15	US-10-154-884B-9051
27	67	11.2	572	9	US-09-864-761-9416
28	66.6	11.2	328	9	US-09-786-692-5737
29	66.6	11.2	328	14	US-10-040-862-5737
30	66.6	11.2	328	15	US-10-057-475B-5737
31	66.6	11.2	328	15	US-10-154-884B-5737
32	63	10.6	476	10	US-09-918-995-32385
33	60	10.1	879	14	US-10-157-031-13
34	57	9.5	445	15	US-10-378-029-25
35	54	9.0	2172	9	US-09-966-880A-15
36	53.2	8.9	821	15	US-10-388-934-189
37	50.2	8.4	674	14	US-10-257-826A-150
38	49.4	8.3	556	9	US-09-796-692-5061
39	49.4	8.3	556	14	US-10-040-862-5061
40	49.4	8.3	556	15	US-10-057-475B-5061
41	49.4	8.3	556	15	US-10-154-884B-5061
42	48	8.0	553	9	US-09-796-692-4504
43	48	8.0	553	14	US-10-040-862-4504
44	48	8.0	553	15	US-10-057-475B-4504
45	48	8.0	553	15	US-10-154-884B-4504

#### ALIGNMENTS

RESULT 1  
US-09-966-880A-7  
Sequence 7, Application US/09966880A  
Patent No. US20020164743A1  
GENERAL INFORMATION:  
APPLICANT: Honjo, Taku  
TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
FILE REFERENCE: 06501-088001  
CURRENT APPLICATION NUMBER: US/09/966, 880A  
CURRENT FILING DATE: 2001-09-28  
PRIOR APPLICATION NUMBER: PCT/JP00/01918  
PRIOR FILING DATE: 2000-03-28  
PRIOR APPLICATION NUMBER: JP 11-371382  
PRIOR FILING DATE: 1999-12-27  
PRIOR APPLICATION NUMBER: JP 11-178999  
PRIOR FILING DATE: 1999-06-24  
PRIOR APPLICATION NUMBER: JP 11-87192  
PRIOR FILING DATE: 1999-03-29  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 7  
LENGTH: 2818  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (80)...(673)  
FEATURE:  
NAME/KEY: 5'UTR  
LOCATION: (1)...(79)  
FEATURE:  
NAME/KEY: 3'UTR  
LOCATION: (677)...(2818)  
US-09-966-880A-7  
Query Match 100.0%; Score 597; DB 9; Length 2818;  
Best Local Similarity 100.0%; Pred. No. 1.4e-187;

Sequence 112023,  
Sequence 112024,  
Sequence 1099, Ap  
Sequence 30237, A  
Sequence 13472, A  
Sequence 1759, A  
Sequence 25187, A  
Sequence 9051, Ap  
Sequence 9051, Ap  
Sequence 9051, Ap  
Sequence 9051, Ap  
Sequence 9416, Ap  
Sequence 5737, Ap  
Sequence 5737, Ap  
Sequence 5737, Ap  
Sequence 5737, Ap  
Sequence 32385, A  
Sequence 13, Appl  
Sequence 25, Appl  
Sequence 15, Appl  
Sequence 189, Appl  
Sequence 150, Ap  
Sequence 5061, Ap  
Sequence 5061, Ap  
Sequence 5061, Ap  
Sequence 5061, Ap  
Sequence 4504, Ap  
Sequence 4504, Ap  
Sequence 4504, Ap  
Sequence 4504, Ap



Matches 597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ATGACAGCCTCTGATGAAACCGAGAGATTCTTTACCAATTCAAAATGTCGCGTGG 60
DB 80 ATGACAGCCTCTGATGAAACCGAGAGATTCTTTACCAATTCAAAATGTCGCGTGG 139
QY 61 GCTAAGGAGTGGCGTGAAGACCTACCTGTGTAAGTAAGAGCGGTGACAGTGTACA 120
DB 140 GCTAAGGAGTGGCGTGAAGACCTACCTGTGTAAGTAAGAGCGGTGACAGTGTACA 199
QY 121 TCCCTTTCACTGAGCTTTGTTATCTTCGCAATTAAGAGCGCTCCACCTGAAATTCCTC 180
DB 200 TCCCTTTCACTGAGCTTTGTTATCTTCGCAATTAAGAGCGCTCCACCTGAAATTCCTC 259
QY 181 TTCCCTCCGCTACATCTCGGACCTGAGACCTGAGCCGCGCGCTGACACCGCTGACCTGG 240
DB 260 TTCCCTCCGCTACATCTCGGACCTGAGACCTGAGCCGCGCGCTGACACCGCTGACCTGG 319
QY 241 TTCACTCTGAGACCCCTCTGACGACTGTGCGGACATGTGGCCGACATTTCTGCGAGGG 300
DB 320 TTCACTCTGAGACCCCTCTGACGACTGTGCGGACATGTGGCCGACATTTCTGCGAGGG 379
QY 301 AACCCCAACCTCACTGCTGAGGATCTTCAACGCGCGCTCTACTTCTGTGAGGACCGCAAG 360
DB 380 AACCCCAACCTCACTGCTGAGGATCTTCAACGCGCGCTCTACTTCTGTGAGGACCGCAAG 439
QY 361 GCTGAGCCGAGGGGCTGCGCGCGCTGACACCGCGCGGGGTGCAAAATAGCCATATGACC 420
DB 440 GCTGAGCCGAGGGGCTGCGCGCGCTGACACCGCGCGGGGTGCAAAATAGCCATATGACC 499
QY 421 TTCAAGATATTTTATCTGCTGAGAACTTTGTGAGAAAACCATGAAGAACTTTCAAA 480
DB 500 TTCAAGATATTTTATCTGCTGAGAACTTTGTGAGAAAACCATGAAGAACTTTCAAA 559
QY 481 GCCTGGAGAGGGCTGCAATGAATAATTCAGTTCCTCTCCAGACAGCTTCGCGCATCTCT 540
DB 560 GCCTGGAGAGGGCTGCAATGAATAATTCAGTTCCTCTCCAGACAGCTTCGCGCATCTCT 619
QY 541 TTGCCCTGTATGAGGTGATGACTTACGAGACGCAATTTGTAATCTTTGGACCTTTGA 597
DB 620 TTGCCCTGTATGAGGTGATGACTTACGAGACGCAATTTGTAATCTTTGGACCTTTGA 676

RESULT 2
US-09-966-880a-1
; Sequence 1, Application US/09966880A
; Patent No. US2002016473A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; APPLICANT: Muramatsu, Masamichi
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966, 880A
; PRIORITY FILING DATE: 2001-09-28
; PRIORITY APPLICATION NUMBER: PCT/JP00/01918
; PRIORITY FILING DATE: 2000-03-28
; PRIORITY APPLICATION NUMBER: JP 11-371382
; PRIORITY FILING DATE: 1999-12-27
; PRIORITY APPLICATION NUMBER: JP 11-178999
; PRIORITY FILING DATE: 1999-06-24
; PRIORITY APPLICATION NUMBER: JP 11-87192
; PRIORITY FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2440
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (93)...(686)
; FEATURE:
; NAME/KEY: 5'UTR
```

```
; LOCATION: (1)...(92)
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: (690)...(2440)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(2440)
; OTHER INFORMATION: n = A,T,C or G
US-09-966-880a-1
```

Query Match 79.6%; Score 475.4; DB 9; Length 2440;  
Best Local Similarity 87.3%; Pred. No. 4e-147;  
Matches 521; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

```
QY 1 ATGACAGCCTCTGATGAAACCGAGAGATTCTTTACCAATTCAAAATGTCGCGTGG 60
DB 93 ATGACAGCCTCTGATGAAACCGAGAGATTCTTTACCAATTCAAAATGTCGCGTGG 152
QY 61 GCTAAGGAGTGGCGTGAAGACCTACCTGTGTAAGTAAGAGCGGTGACAGTGTACA 120
DB 153 GCTAAGGAGTGGCGTGAAGACCTACCTGTGTAAGTAAGAGCGGTGACAGTGTACA 212
QY 121 TCCCTTTCACTGAGCTTTGTTATCTTCGCAATTAAGAGCGCTCCACCTGAAATTCCTC 180
DB 213 TCCCTTTCACTGAGCTTTGTTATCTTCGCAATTAAGAGCGCTCCACCTGAAATTCCTC 272
QY 181 TTCCCTCCGCTACATCTCGGACCTGAGACCTGAGCCGCGCGCTGACACCGCTGACCTGG 240
DB 273 TTCCCTCCGCTACATCTCGGACCTGAGACCTGAGCCGCGCGCTGACACCGCTGACCTGG 332
QY 241 TTCACTCTGAGACCCCTCTGACGACTGTGCGGACATGTGGCCGACATTTCTGCGAGGG 300
DB 333 TTCACTCTGAGACCCCTCTGACGACTGTGCGGACATGTGGCCGACATTTCTGCGAGGG 392
QY 301 AACCCCAACCTCACTGCTGAGGATCTTCAACGCGCGCTCTACTTCTGTGAGGACCGCAAG 360
DB 393 AACCCCAACCTCACTGCTGAGGATCTTCAACGCGCGCTCTACTTCTGTGAGGACCGCAAG 452
QY 361 GCTGAGCCGAGGGGCTGCGCGCGCTGACACCGCGCGGGGTGCAAAATAGCCATATGACC 420
DB 453 GCTGAGCCGAGGGGCTGCGCGCGCTGACACCGCGCGGGGTGCAAAATAGCCATATGACC 512
QY 421 TTCAAGATATTTTATCTGCTGAGAACTTTGTGAGAAAACCATGAAGAACTTTCAAA 480
DB 513 TTCAAGATATTTTATCTGCTGAGAACTTTGTGAGAAAACCATGAAGAACTTTCAAA 572
QY 481 GCCTGGAGAGGGCTGCAATGAATAATTCAGTTCCTCTCCAGACAGCTTCGCGCATCTCT 540
DB 573 GCCTGGAGAGGGCTGCAATGAATAATTCAGTTCCTCTCCAGACAGCTTCGCGCATCTCT 632
QY 541 TTGCCCTGTATGAGGTGATGACTTACGAGACGCAATTTGTAATCTTTGGACCTTTGA 597
DB 633 TTGCCCTGTATGAGGTGATGACTTACGAGACGCAATTTGTAATCTTTGGACCTTTGA 689

RESULT 3
US-09-966-880a-10
; Sequence 10, Application US/09966880A
; Patent No. US2002016473A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; APPLICANT: Muramatsu, Masamichi
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966, 880A
; PRIORITY FILING DATE: 2001-09-28
; PRIORITY APPLICATION NUMBER: PCT/JP00/01918
; PRIORITY FILING DATE: 2000-03-28
; PRIORITY APPLICATION NUMBER: JP 11-371382
; PRIORITY FILING DATE: 1999-12-27
; PRIORITY APPLICATION NUMBER: JP 11-178999
; PRIORITY FILING DATE: 1999-06-24
; PRIORITY APPLICATION NUMBER: JP 11-87192
```



;; PRIOR FILING DATE: 1999-03-29  
;; NUMBER OF SEQ ID NOS: 36  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 10  
;; LENGTH: 6564  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
US-09-966-880A-10

Query Match 45.8%; Score 273.4; DB 9; Length 6564;  
Best Local Similarity 99.6%; Pred. No. 1.3e-79;  
Matches 274; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 155 AGAAGCGCTGCGACGAGTAATGCTTCTCCGCTACATCTGGAGCTGGAGACCTAGACC 214  
DB 2590 AGAAGCGCTGCGACGAGTAATGCTTCTCCGCTACATCTGGAGCTGGAGACCTAGACC 2649  
QY 215 CTGGCCGCTGTCTACCGGCTCACTGTTCACCTCTCTGGAGCCCTGCTACGACTGTGCC 274  
DB 2650 CTGGCCGCTGTCTACCGGCTCACTGTTCACCTCTCTGGAGCCCTGCTACGACTGTGCC 2709  
QY 275 GACATGTGGCCGACTTCTGTGAGAGGAAACCCCACTCACTTGAAGATCTTACCGCGC 334  
DB 2710 GACATGTGGCCGACTTCTGTGAGAGGAAACCCCACTCACTTGAAGATCTTACCGCGC 2769  
QY 335 GCCTCTACTTCTGTGAGAGGAAACCCCACTCACTTGAAGATCTTACCGCGC 394  
DB 2770 GCCTCTACTTCTGTGAGAGGAAACCCCACTCACTTGAAGATCTTACCGCGC 2829  
QY 395 CCGGGGTGCAAAATAGCCATCATGACCTTCAAGAT 429  
DB 2830 CCGGGGTGCAAAATAGCCATCATGACCTTCAAGAT 2864

## RESULT 4

US-09-966-880A-35  
;; Sequence 35, Application US/09966880A  
;; Patent No. US20020164743A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Muramatsu, Masamichi  
;; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
;; FILE REFERENCE: 06501-088001  
;; CURRENT APPLICATION NUMBER: US/09/966,880A  
;; PRIOR APPLICATION NUMBER: PCT/JP00/01918  
;; PRIOR FILING DATE: 2000-03-28  
;; PRIOR APPLICATION NUMBER: JP 11-371382  
;; PRIOR FILING DATE: 1999-12-27  
;; PRIOR APPLICATION NUMBER: JP 11-178999  
;; PRIOR FILING DATE: 1999-06-24  
;; PRIOR APPLICATION NUMBER: JP 11-87192  
;; PRIOR FILING DATE: 1999-03-29  
;; NUMBER OF SEQ ID NOS: 36  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 35  
;; LENGTH: 11204  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
US-09-966-880A-35

Query Match 45.8%; Score 273.4; DB 9; Length 11204;  
Best Local Similarity 99.6%; Pred. No. 1.8e-79;  
Matches 274; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 155 AGAAGCGCTGCGACGAGTAATGCTTCTCCGCTACATCTGGAGCTGGAGACCTAGACC 214  
DB 7805 AGAAGCGCTGCGACGAGTAATGCTTCTCCGCTACATCTGGAGCTGGAGACCTAGACC 7864  
QY 215 CTGGCCGCTGTCTACCGGCTCACTGTTCACCTCTCTGGAGCCCTGCTACGACTGTGCC 274  
DB 7865 CTGGCCGCTGTCTACCGGCTCACTGTTCACCTCTCTGGAGCCCTGCTACGACTGTGCC 7924

QY 275 GACATGTGGCCGACTTCTGTGAGAGGAAACCCCACTCACTTGAAGATCTTACCGCGC 334  
DB 7925 GACATGTGGCCGACTTCTGTGAGAGGAAACCCCACTCACTTGAAGATCTTACCGCGC 7984  
QY 335 GCCTCTACTTCTGTGAGAGGAAACCCCACTCACTTGAAGATCTTACCGCGC 394  
DB 7985 GCCTCTACTTCTGTGAGAGGAAACCCCACTCACTTGAAGATCTTACCGCGC 8044  
QY 395 CCGGGGTGCAAAATAGCCATCATGACCTTCAAGAT 429  
DB 8045 CCGGGGTGCAAAATAGCCATCATGACCTTCAAGAT 8079

## RESULT 5

US-09-966-880A-13  
;; Sequence 13, Application US/09966880A  
;; Patent No. US20020164743A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Honjo, Tasuku  
;; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
;; FILE REFERENCE: 06501-088001  
;; CURRENT APPLICATION NUMBER: US/09/966,880A  
;; PRIOR APPLICATION NUMBER: PCT/JP00/01918  
;; PRIOR FILING DATE: 2000-03-28  
;; PRIOR APPLICATION NUMBER: JP 11-371382  
;; PRIOR FILING DATE: 1999-12-27  
;; PRIOR APPLICATION NUMBER: JP 11-178999  
;; PRIOR FILING DATE: 1999-06-24  
;; PRIOR APPLICATION NUMBER: JP 11-87192  
;; PRIOR FILING DATE: 1999-03-29  
;; NUMBER OF SEQ ID NOS: 36  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 13  
;; LENGTH: 271  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
US-09-966-880A-13

Query Match 45.4%; Score 271; DB 9; Length 271;  
Best Local Similarity 100.0%; Pred. No. 1.4e-79;  
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 157 AACGGCTGCAAGTGAATGCTTCTCCGCTACATCTGGAGCTGGAGACCTAGACC 216  
DB 1 AACGGCTGCAAGTGAATGCTTCTCCGCTACATCTGGAGCTGGAGACCTAGACC 60  
QY 217 GCGCGCTGTCTACCGGCTCACTGTTCACCTCTCTGGAGCCCTGCTACGACTGTGCC 276  
DB 61 GCGCGCTGTCTACCGGCTCACTGTTCACCTCTCTGGAGCCCTGCTACGACTGTGCC 120  
QY 277 CATGTGGCCGACTTCTGTGAGAGGAAACCCCACTCACTTGAAGATCTTACCGCGC 336  
DB 121 CATGTGGCCGACTTCTGTGAGAGGAAACCCCACTCACTTGAAGATCTTACCGCGC 180  
QY 337 CTCTACTTCTGTGAGAGGAAACCCCACTCACTTGAAGATCTTACCGCGC 396  
DB 181 CTCTACTTCTGTGAGAGGAAACCCCACTCACTTGAAGATCTTACCGCGC 240  
QY 397 GGGGTGCAAAATAGCCATCATGACCTTCAAG 427  
DB 241 GGGGTGCAAAATAGCCATCATGACCTTCAAG 271

RESULT 6  
US-09-966-880A-12  
;; Sequence 12, Application US/09966880A  
;; Patent No. US20020164743A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Honjo, Tasuku  
;; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE



FILE REFERENCE: 06501-088001  
CURRENT APPLICATION NUMBER: US/09/966, 880A  
CURRENT FILING DATE: 2001-09-28  
PRIOR APPLICATION NUMBER: PCT/JP00/01918  
PRIOR FILING DATE: 2000-03-28  
PRIOR APPLICATION NUMBER: JP 11-371382  
PRIOR FILING DATE: 1999-12-27  
PRIOR APPLICATION NUMBER: JP 11-178999  
PRIOR FILING DATE: 1999-06-24  
PRIOR APPLICATION NUMBER: JP 11-87192  
PRIOR FILING DATE: 1999-03-29  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 12  
LENGTH: 148  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-966-880A-12

Query Match 24.8%; Score 148; DB 9; Length 148;  
Best Local Similarity 100.0%; Pred. No. 9, 6e-39; Indels 0; Gaps 0;  
Matches 148; Conservative 0; Mismatches 0;

QY 9 CCTCTGATGACCGGAGAGTTTCTTACCAATTCAAAATGTCGGCTGGCTAAGG 68  
DB 1 CCTTTATGACCGGAGAGTTTCTTACCAATTCAAAATGTCGGCTGGCTAAGG 60  
QY 69 TCGGCGTGAACCTACTGCTGCTAGTGAAGGCGTGAAGCTGCTAATCTTTTC 128  
DB 61 TCGGCGTGAACCTACTGCTGCTAGTGAAGGCGTGAAGCTGCTAATCTTTTC 120  
QY 129 ACTGACTTGTGTTTCTTGGCAATAG 156  
DB 121 ACTGACTTGTGTTTCTTGGCAATAG 148

## RESULT 7

US-10-460-923-1  
Sequence 1, Application US/10460923  
Publication No. US2004000951A1  
GENERAL INFORMATION:  
APPLICANT: MALIM, Michael H.  
APPLICANT: SHEEHY, Ann M.  
APPLICANT: HARRIS, Reuben S.  
APPLICANT: BISHOP, Kate N.  
APPLICANT: NEUBERGER, Michael S.  
APPLICANT: GADDIS, Nathan C.  
APPLICANT: SIMON, James H.M.  
TITLE OF INVENTION: DNA Deamination Mediates Innate Immunity to Retroviral Infection  
FILE REFERENCE: 2253-74380  
CURRENT APPLICATION NUMBER: US/10/460, 923  
CURRENT FILING DATE: 2003-06-13  
PRIOR APPLICATION NUMBER: US 60/388, 513  
PRIOR FILING DATE: 2002-06-13  
PRIOR APPLICATION NUMBER: US 60/472, 952  
PRIOR FILING DATE: 2003-05-23  
NUMBER OF SEQ ID NOS: 12  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 1  
LENGTH: 1155  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-460-923-1

Query Match 19.8%; Score 118.4; DB 15; Length 1155;  
Best Local Similarity 58.9%; Pred. No. 2, 2e-28; Indels 6; Gaps 1;  
Matches 225; Conservative 0; Mismatches 151;

QY 158 ACGGTCGACAGTATGCTTCTCCGCTACATCTGAGTGGAGACTAGACCTG 217  
DB 761 AAGGCGCATGACAGCTGCTTCTCGAAGGATTCCTTTGGAAGCTGAGCTGG 820  
QY 218 GCGGCTGCTACCGGCTACCTGCTTCACTCTCTGAGAGCCCTGCTAGAGCTGAGCCGAC 277

DB 821 ACCAGGACTACAGGTTACCTGCTCACTCTGAGAGCCCTCTTACGCTGAGCCAG 880  
QY 278 ATGTCGCGACTTCTGCGAGGAGACCCCACTCAGTCTGAGATCTTACCGCGCC 337  
DB 861 AATGCGTAATTTATTTCAAAAACAAACAGTGAAGCTGCTGATCTTACGCGCCGA 940  
QY 338 TCTACTTCTGAGGACCGAAGGCTAGGCGGAGGCTGCGGCTGACCGCGCCG 397  
DB 941 TCTA-----TGATGATCAAGAAAGATGTCAGAGGCGGCTGCGGACCTGCGAGGCTG 994  
QY 398 GGGTGAATTAAGCATATGACTTCAAGATTTATTTACTGTGAAATCTTTTAG 457  
DB 995 GGGCCAAATTTCAATATGATGATGATGATGATGATGATGATGATGATGATGATG 1054  
QY 458 AAACCATGAAAGAACTTTAAAGCTGGGAGGCGTGAAGAAATTCAGTGTCTCT 517  
DB 1055 ACCACGAGGATGTCCTTCCAGGCTGGATGAGTGAAGACAGCCAGACCTGA 1114  
QY 518 CCAAGAGCTTGGGCGATCT 539  
DB 1115 GTGGAGGCTGCGGCAATCT 1136

## RESULT 8

US-09-729-674-173  
Sequence 173, Application US/09729674  
Patent No. US2001003935A1  
GENERAL INFORMATION:  
APPLICANT: JACOBS, Kenneth  
APPLICANT: MCCOY, John M.  
APPLICANT: LAVELLE, Edward R.  
APPLICANT: COLLINS-RACIE, Lisa A.  
APPLICANT: EVANS, Cheryl  
APPLICANT: WETBERG, David  
APPLICANT: TREACY, Maurice  
APPLICANT: AGOSTINO, Michael J.  
APPLICANT: STEINGER, Robert J.  
APPLICANT: SPALDING, Vikki  
APPLICANT: WONG, Gordon G.  
APPLICANT: CLARK, Hilary  
APPLICANT: FECHTEL, Kim  
TITLE OF INVENTION: GENETICS INSTITUTE, INC.  
FILE REFERENCE: 6055-64X  
CURRENT APPLICATION NUMBER: US/09/729, 674  
CURRENT FILING DATE: 2000-12-04  
PRIOR APPLICATION NUMBER: 09/539, 330  
PRIOR FILING DATE: 2000-03-30  
NUMBER OF SEQ ID NOS: 283  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 173  
LENGTH: 1534  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-729-674-173

Query Match 19.8%; Score 118.4; DB 9; Length 1534;  
Best Local Similarity 58.9%; Pred. No. 2, 5e-28; Indels 6; Gaps 1;  
Matches 225; Conservative 0; Mismatches 151;

QY 158 ACGGTCGACAGTATGCTTCTCCGCTACATCTGAGTGGAGACTAGACCTG 217  
DB 811 AAGGCGCATGACAGCTGCTTCTCGAAGGATTCCTTTGGAAGCTGAGCTGG 870  
QY 218 GCGGCTGCTACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 277  
DB 871 ACCAGGACTACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 930  
QY 278 ATGTCGCGACTTCTGCGAGGAGACCCCACTCAGTCTGAGATCTTACCGCGCC 337  
DB 931 AATGCGTAATTTATTTCAAAAACAAACAGTGAAGCTGCTGAGCTGAGCTGAGCCGCA 990



Accession	Sequence	Position
QY	TCATCTTGTGAGGACCGCAAGGCTGAGCCCGAAGGGCTGGGCGGTGACCGCGCG	397
QY	TCATA-----TGATATCAAGGAAGATGTCAAGAGGGCTGTCCACCTGTGCGTAGGCTG	1044
Db		
QY	GGGTGCAATAGCCATCATGACCTTCAAGATTATTTTACTGCTGGAATACTTTGTAG	457
Db	GGGCGAAATTTTCATTAATGACATCATGATGATTTAAACACATGCTGGGACACCTTTGTGG	1104
QY	AAATCCATGAAAGAACTTTCAAGGCTGGGAAGGCTGCATGAAATTCAGTTGCTCTCT	517
Db	ACCAACAGAGGATGTCCCTTCCAGCCCTTGGAATGACTAGATGAGCACAGCCAAAGACTGA	1164
QY	CCAGACAGCTTGGCGGCATCCT	539
Db	GTGGGAGGCTGGCGGCATTTCT	1186

## RESULT 9

```

Sequence 699, Application US/09925300
Patent No. US20020151681A1
GENERAL INFORMATION:
APPLICANT: Craig Rosen,
APPLICANT: Steve Ruben
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA101
CURRENT APPLICATION NUMBER: US/09/925,300
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05988
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1890
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 699
LENGTH: 987
TYPE: DNA
ORGANISM: Homo sapiens
US-09-925-300-699

```

Query Match 19.2%; Score 114.8; DB 9; Length 987;

Matches 289; Conservative 0; Mismatches 232; Indels 15; Gaps 2;

Oy 16 ATGAAACCGAGAAAGTTTCTTTACCAATTCAAAAATGTCGCTGGAGCTAAAGGGTGGGGGT 75  
 Db 132 ATGATTCAGGACAACTTCTACTCTCAATTTAAAAAATTATGGGAAACCAACATGTGGAAAC 191  
 Oy 76 GAGACCTTACCTGTGTCTACGTATAGTAAGAGCGGTGACAGTGCATACCTCTTTCACCTGAC 135  
 Db 192 GAACTTGTGCTGTGTCTTACCGGTGAAGGTATAAAGCGCGCTCAGTTGTCTCTCGAAG 251  
 Oy 136 TTTGGTATCTTCGCAA-----TAAGAACGGGTGCGCAGTGGATTTGCTCTTC 183  
 Db 252 ACGGGCGCTTCCGAAACCGAGTGGATTCTGAGACCATTTGATGCAAGAAAGGTGCTTC 311  
 Oy 184 CTCGCGCTACATCTCGGACTGGAGACCTAGACCTGTGCGGTGCTACCGGCTCACCTGTTC 243  
 Db 312 CTCTCTTGGTTCTGGAGGACATACTGTCTCTTAACCAAATACCAAGTCACTGTGTAC 371  
 Oy 244 AACTTCCTGAGGCCCTGTGCTACGACTGTGCCCCGACATGTGGCCGCACTTTCGTGGAGGAAC 303  
 Db 372 ACATTTTGAAGCCCTTGGCCACACTGTGTGAAGGAGATGGCCGAGATTCTGTGCCAGGAC 431  
 Oy 304 CCCAACCTCAGTCTGAGGATTTTCAACCGCGCGCTTACTTCTGTGAAGACCGCAAGGT 363  
 Db 432 AGCAACGTGAATCTACCATCTTCAACGGCGCGCTCTACTTACCTTCCAGATC---CAGGT 488  
 Oy 364 GAGCCCGAGGGGCTGCGGCGGCTGCACCGCGCCGGGGTGCAAATAGCCATCATGACCTTC 423  
 Db 489 TACCAAGAGGGGCTCCGACGCTTGATGTCAGAAAGGGGTGCTGTGGAGATCATGAGATAT 548  
 Oy 424 AAAGATTATTTTACTGCTGGAAATCTTTTGTAGAAAAACATGAAAGAACTTTCAAAGCC 483

Db 549 GAAAGATTTTAAATATTTGGAAAACTTTGTGTCAATGTATTAATGAGCCATTAAAGCTT 608

Gy 484 TGGGAAGGCTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCTT 539

Db 609 TGGAAAGGATTAATAAACCAACTTTGCACCTTCGAAAAAGAGCTTACGGAGAGCTT 664

RESULT 10

```

: Sequence 268, Application US/10098841
: Publication No. US20020197679A1
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Asundi, Yindod
APPLICANT: Xu, Chongjun
APPLICANT: Zhou, Ping
APPLICANT: Ma, Yundong
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhao, Qing A.
APPLICANT: Ren, Feiyan
APPLICANT: Chen, Rui-hong
APPLICANT: Wang, Dunlul
APPLICANT: Wang, Zhiwei
APPLICANT: Wehrman, Tom
APPLICANT: Zheng, Jie
APPLICANT: Qian, Xiaohong B.
APPLICANT: Drmanec, Radej T.
TITLE OF INVENTION: No. US20020197679A1el Nucleic Acids and
TITLE OF INVENTION: Polypeptides
FILE REFERENCE: 784CIP2
CURRENT APPLICATION NUMBER: US/10/098, 841
CURRENT FILING DATE: 2002-03-13
PRIOR APPLICATION NUMBER: 09/598,042
PRIOR FILING DATE: 2000-06-20
PRIOR APPLICATION NUMBER: 09/552,317
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/488,725
PRIOR FILING DATE: 2000-01-21
NUMBER OF SEQ ID NOS: 331
SOFTWARE: pc_Fl_genes Version 1.0
SEQ ID NO 268
LENGTH: 1143
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (128)..(700)
US-10-098-841-268

```

Query Match	Score	DB	Length
19.28;	114.8;	13;	1143;

Matches 289; Conservative 0; Mismatches 232; Indels 15; Gaps 2;

QY 16 ATGAACCGAGGAGAGTTTCTTACCAATTCAAAAATGCGCTGGCTACAAAGGTCGGCGT 75  
 Db 161 ATGTATCCAGGCACATTTCTACTTCCATTTAAATTAACCTATGGAAAGCCACGATCCGAAAC 22  
 QY 76 GAGACCTACCTGTCTGTACTAGTAGAAGAGCGGTGACAGTCACTCTTTTCACTGAC 13  
 Db 221 GAACACTGCTGTGCTTCAACGCTGAAGGTATTAAGACCCGCTCAGTTGTCTCTGGAAG 28  
 QY 136 TTGGTATATCTTGGCAA-----TAAAGACGGCTGCCAGTGGAAATTCCTCTTC 181  
 Db 281 ACGGGCGTCTTCGAAACACAGGTGATTTCTGAGCCCATTTGTCATGACGAAAGGTGCTTC 34  
 QY 184 CTCGCGTACATCTCCGGAACCTGGACCTAGACCCTGGCGGCTCTACCGCGTCACTGGTC 24  
 Db 341 CTCCTTGGTCTGCGACACATACTGTCTCTTAACAAGATACAGGTCACTTGATAC 40  
 QY 244 ACCTCTGGAAGCCCTCTGCTACACTGTGCGCGACATGTGGCCGACTTTCTGCGAGGGAAC 30



```

DB 401 ACATTTGAGAGCCCTTCCAGACTTGTGCAAGGAGGTTGCCGAGTTCTTCCAGGAC 460
QY 304 CCCAACCTCAGTCTGAGATCTTTACCGCGGCTCTTACTTTGTGAGAGCCGAGAGCT 363
DB 461 AGCAACGTAACCTACCACTCTTACCGCCGCTCTACTCTTCACTATC---CATGT 517
QY 364 GAGCCCGAGGGGCTCGCGCGCTGCAACCGCCGCGGGGTGCAAAATGACCATATGACCTTC 423
DB 518 TACCAAGAGGGGCTCCGACGCTGAGTCAAGAGGGGTGCTGTGAGATCATGACATAT 577
QY 424 AAAAGTATTTTACTGCTGAGATCTTTGTGAGAAACATGAAAGAACTTTGAAAGCC 483
DB 578 GAAGATTTTAAATATTTGTTGGGAAAATTGTGTACATGATGAGCCATTCAGCTT 637
QY 484 TGGGAAGGGCTGCATGAAATATTCAGTCTCTCTCCAGACAGCTTCGCGCATCTT 539
DB 638 TGGAAAGGATTAATAAACCACTTTGCACTTCTGAAAAAAGAGCTACGGGAGAGTCT 693

```

## RESULT 11

```

US-09-966-880A-14
Sequence 14, Application US/09966880A
Patent No. US2002018473A1
GENERAL INFORMATION:
APPLICANT: Honjo, Tasuku
APPLICANT: Muramatsu, Masamichi
TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
FILE REFERENCE: 06501-088001
CURRENT APPLICATION NUMBER: US/09/966,880A
PRIOR FILING DATE: 2001-09-28
PRIOR APPLICATION NUMBER: PCT/JP00/01918
PRIOR FILING DATE: 2000-03-28
PRIOR APPLICATION NUMBER: JP 11-371382
PRIOR FILING DATE: 1999-12-27
PRIOR APPLICATION NUMBER: JP 11-178999
PRIOR FILING DATE: 1999-06-24
PRIOR APPLICATION NUMBER: JP 11-87192
PRIOR FILING DATE: 1999-03-29
NUMBER OF SEQ ID NOS: 36
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 14
LENGTH: 116
TYPE: DNA
ORGANISM: Homo sapiens
US-09-966-880A-14

```

```

Query Match 19.2%; Score 114.4; DB 9; Length 116;
Best Local Similarity 99.1%; Pred. No. 1.3e-27;
Matches 115; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 428 ATTATTTTACTGCTGGAATCTTTGTGAGAAACCATGAAAGAACTTTGAAAGCTGG 487
DB 1 ATTATTTTACTGCTGGAATCTTTGTGAGAAACCATGAAAGAACTTTGAAAGCTGG 60
QY 488 AAGGCGTGCATGAAATATTCAGTCTCTCCAGACAGCTTCGCGCATCTTTG 543
DB 61 AAGGCGTGCATGAAATATTCAGTCTCTCCAGACAGCTTCGCGCATCTTTG 116

```

## RESULT 12

```

US-09-880-107-3282
Sequence 3282, Application US/09880107
Patent No. US20020142981A1
GENERAL INFORMATION:
APPLICANT: Horne, Darci T.
APPLICANT: Vockley, Joseph G.
APPLICANT: Scherf, Uwe
APPLICANT: Gene Logic, Inc.
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
FILE REFERENCE: 44921-5028-WO
CURRENT APPLICATION NUMBER: US/09/880,107
CURRENT FILING DATE: 2001-06-14
PRIOR APPLICATION NUMBER: US 60/211,379

```

```

PRIOR FILING DATE: 2000-06-14
PRIOR APPLICATION NUMBER: US 60/237,054
PRIOR FILING DATE: 2000-10-02
NUMBER OF SEQ ID NOS: 3950
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 3282
LENGTH: 1348
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Genbank Accession No. US20020142981A1 U03891
NAME/KEY: unsure
LOCATION: (1)..(1348)
OTHER INFORMATION: n = a or c or g or t
US-09-880-107-3282

```

```

Query Match 18.5%; Score 110.4; DB 9; Length 1348;
Best Local Similarity 59.3%; Pred. No. 1.1e-25;
Matches 230; Conservative 0; Mismatches 146; Indels 12; Gaps 2;

```

```

QY 158 ACGGCTGCCACGTGAGATTGCTTCTCCGCTACATCTGAGACTGGAGCTAGACCTTG 217
DB 275 ACGGCGGCATGCGGAGCTGCGCTTCTGGAACCTGGTCTTCTTGACAGTTGGACCCGG 334
QY 218 GCGGCTGCTACCGCGCTACCTGCTTCACTCTCTGAGCCCGCTAC-----GACTGTG 271
DB 335 CCCAGATCTACAGGGTCACTTGCTTCATCTCTGAGCCCTGCTTCTCGGGCTGTG 394
QY 272 CCGGACATGCGCGGCTTCTGCGAGGAAACCCCACTCACTGAGAGATCTTACCG 331
DB 395 CCGGAGATGCTGCTGCTTCTTCAAGAAACACACACTGAGACTGGGCATCTTGGCTG 454
QY 332 CCGGCTCTACTTCTGTGAGAACCGCAAGCTGAGCGCGCTGCGCGCTGACCC 391
DB 455 CCGGATCTA-----TGATTAAGACCCCTATATAAGAGCGCTGCAATGCTGCGGG 508
QY 392 GCGCGGGGCTGCAAAATAGCCATATGACTTCAAGATTTTATCTGCTGGAATCTT 451
DB 509 ATGCTGGGCGCCAAAGCTCCATCATGACCTACGATAAATTTAAGCATCTGAGACCT 568
QY 452 TTGTAGAAACCATGAAAGAACTTCAAGCCCTGGAGAGGCTGCAAAATTCAGTTC 511
DB 569 TTGTAGACCAACGAGGATGTCCTTCAAGCCCTGGAGATGATGATGAGCAGACCAAG 628
QY 512 GTCTCTCCAGACAGCTTCGCGCATCTT 539
DB 629 CCTGAGTGGAGAGCTTCGCGGCATTTCT 656

```

## RESULT 13

```

US-10-247-671-62
Sequence 62, Application US/10247671
Publication No. US20030194721A1
GENERAL INFORMATION:
APPLICANT: Mikita, Thomas
APPLICANT: Shiffman, Dov
APPLICANT: Porter, Gordon, J.
APPLICANT: Kaser, Matthew R.
TITLE OF INVENTION: GENES EXPRESSED IN TREATED FOAM CELLS
FILE REFERENCE: PA-0050 US
CURRENT APPLICATION NUMBER: US/10/247,671
CURRENT FILING DATE: 2002-09-18
PRIOR APPLICATION NUMBER: 60/323,784
PRIOR FILING DATE: 2001-09-19
NUMBER OF SEQ ID NOS: 186
SOFTWARE: PERL Program
SEQ ID NO 62
LENGTH: 823
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. US20030194721A1 135626CHI

```







Mon Mar 15 09:28:01 2004

us-09-966-880a-7\_copy\_80\_676.rnpb

Page 8

QY 458 AAAACCATGAAAGAACTTCAAGCGTGGGAAAGGCTGCATGAAATTCAGTTGATCTCT 517

DB 1742 ACCACCAAGGATGTCCTCTTCCAGCCTGGATGATCATGATAGACAAGCCAAAGCTGA 1801

QY 518 CCAGACAGCTTCGGCGCATCTCT 539

DB 1802 GTGGAGAGCTGCGCGCATTCCT 1823

Search completed: March 12, 2004, 12:26:29  
Job time : 250 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: March 11, 2004, 17:34:19 ; Search time 22 Seconds  
(without alignments)  
464.634 Million cell updates/sec

Title: US-09-966-880A-8  
Perfect score: 1086  
Sequence: 1 MDSLMNRRKFLYQFNVRW.....ILPLFYVDDLRDAPRTLGI 198

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum March 0%  
Maximum March 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PTCUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	369.5	34.0	190	1	US-08-816-241-1 Sequence 1, Appl
2	369.5	34.0	190	3	US-09-128-395-1 Sequence 1, Appl
3	271	25.0	222	1	US-08-687-895-1 Sequence 1, Appl
4	271	25.0	222	2	US-09-040-482-1 Sequence 1, Appl
5	245	22.6	116	1	US-08-687-895-3 Sequence 3, Appl
6	245	22.6	116	1	US-08-816-241-3 Sequence 3, Appl
7	245	22.6	116	2	US-09-040-482-3 Sequence 3, Appl
8	245	22.6	116	3	US-09-128-395-3 Sequence 3, Appl
9	225	20.7	275	4	US-09-079-030-224 Sequence 2, Appl
10	216.5	19.9	236	1	US-08-687-895-4 Sequence 4, Appl
11	216.5	19.9	236	1	US-08-816-241-4 Sequence 4, Appl
12	216.5	19.9	236	2	US-09-040-482-4 Sequence 4, Appl
13	216.5	19.9	236	2	US-09-128-395-4 Sequence 4, Appl
14	210	19.3	236	1	US-08-158-682A-4 Sequence 4, Appl
15	198	18.2	229	1	US-08-158-682A-2 Sequence 2, Appl
16	198	18.2	229	1	US-08-015-203-2 Sequence 2, Appl
17	198	18.2	229	1	US-08-687-895-5 Sequence 5, Appl
18	198	18.2	229	1	US-08-816-241-5 Sequence 5, Appl
19	198	18.2	229	2	US-09-040-482-5 Sequence 5, Appl
20	198	18.2	229	2	US-09-128-395-5 Sequence 5, Appl
21	81	7.3	440	4	US-09-092-315-3 Sequence 3, Appl
22	81	7.5	440	4	US-09-733-524A-3 Sequence 3, Appl
23	80.5	7.4	451	4	US-09-107-532A-6444 Sequence 6444, Ap
24	80	7.4	476	4	US-09-092-315-5 Sequence 5, Appl
25	79.5	7.3	476	4	US-09-733-524A-5 Sequence 5, Appl
26	77.5	7.1	402	6	US-09-252-991A-24319 Sequence 24319, A
27					Patent No. 5352575

28	75.5	7.0	343	1	US-08-180-209B-56 Sequence 56, Appl
29	75.5	7.0	343	3	US-08-474-853-56 Sequence 56, Appl
30	75.5	7.0	343	3	US-09-166-205B-56 Sequence 56, Appl
31	75.5	7.0	343	5	PCR-US94-02629-56 Sequence 5, Appl
32	75.5	7.0	347	4	US-09-806-658-10 Sequence 10, Appl
33	73.5	6.8	533	1	US-08-488-305A-6 Sequence 6, Appl
34	72.5	6.7	366	6	5470718-4 Patent No. 5470718
35	72	6.6	487	4	US-09-489-039A-9810 Sequence 9810, Ap
36	72	6.6	1094	2	US-08-680-326-40 Sequence 40, Appl
37	72	6.6	1400	1	US-08-080-255-7 Sequence 7, Appl
38	72	6.6	1400	3	US-08-465-713-7 Sequence 7, Appl
39	72	6.6	1400	5	PCR-US93-05857-7 Sequence 5, Appl
40	72	6.6	3969	3	US-08-061-376-5 Sequence 5, Appl
41	71	6.5	350	4	US-09-489-039A-10384 Sequence 10384, A
42	71	6.5	351	4	US-09-252-991A-23740 Sequence 23740, A
43	71	6.5	860	4	US-09-936-989A-2 Sequence 2, Appl
44	70.5	6.5	925	4	US-09-936-989A-6 Patent No. 5475095
45	70.5	6.5	528	6	5475095-2

## ALIGNMENTS

RESULT 1  
US-08-816-241-1  
Sequence 1, Application US/08816241  
Patent No. 5804185  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Goli, Surya K.  
TITLE OF INVENTION: NOVEL RNA EDITING ENZYME  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESS: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: PASTESQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/816, 241  
FILING DATE: Filed Herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET INFORMATION:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 190 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: PROSTTT09  
CLONE: 1646823  
US-08-816-241-1  
Query Match 34.0%; Score 369.5; DB 1; Length 190;  
Best Local Similarity 44.9%; Pred. No. 2.5e-35;  
Matches 79; Conservative 24; Mismatches 64; Indels 9; Gaps 4;



QY 11 FLYQFNWAKGRRETYLCYVVK--RRDSATSFSLDFGYLRN----KNGCHEVLLFLRY 64  
DB 17 FVFGKMLWEANDRNITWLCFTVEGIKRSVVSMT--GVPRNQVDSHCHAEKCFLSW 74  
QY 65 ISDMWLDPCRCYRTVWFTSWSPCYDCARHVADEFLRGNPNLSLRTFARLYCEBRKAEPE 124  
DB 75 FCDLILSPNKKYQVMTWYTSWSPCPDAGEVAEFLRHSNVNLIIFTALYYFQ-YPCYOE 133  
QY 125 GRLRLHAGVOAINTFDFYCMNTFVENHRTKAEGLHNSVRLSRQLRLL 180  
DB 134 GRLSLSQEGVAEINDYEDFKYCMENFYNDNEPFPKMGKLTNRRLKRLRESL 189

RESULT 2  
US-09-128-395-1  
Sequence 1, Application US/09128395  
Patent No. 6087108  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Goli, Surya K.  
TITLE OF INVENTION: NOVEL RNA EDITTING ENZYME  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/128,395  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/816,241  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0239 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 190 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: PROSTUT09  
CLONE: 1646823  
US-09-128-395-1

Query Match 34.0%; Score 369.5; DB 3; Length 190;  
Best Local Similarity 44.9%; Pred. No. 2,5e-35;  
Matches 79; Conservative 24; Mismatches 64; Indels 9; Gaps 4;

QY 11 FLYQFNWAKGRRETYLCYVVK--RRDSATSFSLDFGYLRN----KNGCHEVLLFLRY 64  
DB 17 FVFGKMLWEANDRNITWLCFTVEGIKRSVVSMT--GVPRNQVDSHCHAEKCFLSW 74  
QY 65 ISDMWLDPCRCYRTVWFTSWSPCYDCARHVADEFLRGNPNLSLRTFARLYCEBRKAEPE 124  
DB 75 FCDLILSPNKKYQVMTWYTSWSPCPDAGEVAEFLRHSNVNLIIFTALYYFQ-YPCYOE 133  
QY 125 GRLRLHAGVOAINTFDFYCMNTFVENHRTKAEGLHNSVRLSRQLRLL 180

DB 134 GRLSLSQEGVAEINDYEDFKYCMENFYNDNEPFPKMGKLTNRRLKRLRESL 189

RESULT 3  
US-08-687-895-1  
Sequence 1, Application US/08687895  
Patent No. 5747319  
GENERAL INFORMATION:  
APPLICANT: Au-Young, Janice  
APPLICANT: Hillman, Phillip R.  
APPLICANT: Hillman, Jennifer L.  
TITLE OF INVENTION: A NOVEL HUMAN MRNA EDITTING ENZYME  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: U.S.  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/687,895  
FILING DATE: Filed Herewith  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0109 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-845-4166  
TELEFAX: 415-855-0555  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 222 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:  
LIBRARY: MUSCNC01  
CLONE: 57953  
US-08-687-895-1

Query Match 25.0%; Score 271; DB 1; Length 222;  
Best Local Similarity 35.8%; Pred. No. 1,1e-23;  
Matches 63; Conservative 25; Mismatches 76; Indels 12; Gaps 5;

QY 11 FLYQFNWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKN--GCHVLLFLRYSDMD 69  
DB 52 FKQFNRVVEYSRGNKTFCLYVVEAQCKGQVQASRGVLEDEAAHAAEAEFNTLLP-A 110  
QY 70 LDPGRCYRTVWFTSWSPCYDCARHVADEFLRGNPNLSLRTFARLYCEBRKAEPE---GL 126  
DB 111 FDPALRYNTVWYVSSPCCACADRIKXTLSKTNLRLLLVGLFWME---EPFIQAL 166  
QY 127 RRLHAGVOAINTFDFYCMNTFVENHRTKAEGLHNSVRLSRQLRRL 179  
DB 167 KKLKACQKRLKPKQDFETWONFVQEBEGSKAFQPMEDIOENFLYBEKLADI 222

RESULT 4  
US-05-040-482-1  
Sequence 1, Application US/09040482  
Patent No. 5916556  
GENERAL INFORMATION:  
APPLICANT: Au-Young, Janice  
APPLICANT: Hillman, Phillip R.  
APPLICANT: Hillman, Jennifer L.  
TITLE OF INVENTION: A NOVEL HUMAN MRNA EDITTING ENZYME



```

; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040,482
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/687,895
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0109 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 222 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY: MUSCNOT1
; CLONE: 57953
; US-09-040-482-1

Query Match      25.0%; Score 271; DB 2; Length 222;
Best Local Similarity 35.8%; Pred. No. 1,1e-23;
Matches 63; Conservative 25; Mismatches 76; Indels 12; Gaps 5;

Cy 11 FLVYKQVNRKAGREERYLCYVGRDSTPSLDFGYLANK-GCHVELLFRYISDMD 69
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 52 FKFGFRVVEYSSGNKTFLCYVVEAQGGQVOASRGYLEDHAAAHAEAFNTILP-A 110
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Cy 70 LDPRCRRVMTFTNSCYOCARIVADFLGNPULSIRITALLYPEDKAPF---GL 126
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 111 FDPALRYNVMTWYSSSCAACADRIXKTLSTKTKRLLLIVGRLFWME---EPEIQAL 166
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Cy 127 RRLRAGVOAIAIMFKDYFCOMTFVENHNE---RTFKAMEGLHENSVALSRQLRRI 179
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 167 KKLKAGCKLRIVKQDFEYVWQNFVEGEGESKAFQPMEDIOENFLYERKLADI 222
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 5
US-08-687-895-3
; Sequence 3, Application US/08687895
; Patent No. 5747319
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: A NOVEL HUMAN MRNA EDITING ENZYME
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,895
; FILING DATE: Filed Herewith
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0109 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 116 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 436941
; US-08-687-895-3

Query Match      22.6%; Score 245; DB 1; Length 116;
Best Local Similarity 46.4%; Pred. No. 5,1e-21;
Matches 51; Conservative 17; Mismatches 32; Indels 10; Gaps 3;

Cy 76 YRVMTFTSWSPCYD-CARHVAADFLRGPNLSLIFPARLYFCBDRKAPF---EGLRLH 130
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 7 YRVMTFTSWSPCYD-CARHVAADFLRGPNLSLIFPARLYFCBDRKAPF---EGLRLH 130
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Cy 131 RAGVOAIAIMFKDYFCOMTFVENHNEHRTFKAMEGLHENSVALSRQLRRI 180
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 62 DAGAAGVSMITVDFEFCYMDTFVYRQGGCFQWDLSEHSQALSGRLAIL 111
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 6
US-08-816-241-3
; Sequence 3, Application US/08816241
; Patent No. 5804185
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL RNA EDITING ENZYME
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/816,241
; FILING DATE: Filed Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0239 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555

```



TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 116 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 436941  
US-09-816-241-3

Query Match 22.6%; Score 245; DB 1; Length 116;  
Best Local Similarity 46.4%; Pred. No. 5.1e-21;  
Matches 51; Conservative 17; Mismatches 32; Indels 10; Gaps 3;

QY 76 YRVWFTSWSPCYD--CARHVDPLRGPNLSLIFARLYFCEDRAEP---EGLRLH 130  
DB 7 YRVWFTSWSPCYD--CARHVDPLRGPNLSLIFARLYFCEDRAEP---EGLRLH 130  
QY 131 RAGVOIATMTFQDYFCNFTFVENHRTFKAMEGLHNSVRLSRQLRIL 180  
DB 62 DAGAQSIVTITDEFYICMDTFVYRQCGCFQPMWGLIEHSQALSGRLRAIL 111

RESULT 7  
US-09-040-482-3  
Sequence 3, Application US/09040482  
Patent No. 5916556  
GENERAL INFORMATION:  
APPLICANT: Au-Young, Janice  
APPLICANT: Hawkins, Phillip R.  
APPLICANT: Hillman, Jennifer L.  
TITLE OF INVENTION: A NOVEL HUMAN RNA EDITING ENZYME  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: U.S.  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/040,482  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/687,895  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0109 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 116 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 436941  
US-09-040-482-3

Query Match 22.6%; Score 245; DB 2; Length 116;

Best Local Similarity 46.4%; Pred. No. 5.1e-21;  
Matches 51; Conservative 17; Mismatches 32; Indels 10; Gaps 3;

QY 76 YRVWFTSWSPCYD--CARHVDPLRGPNLSLIFARLYFCEDRAEP---EGLRLH 130  
DB 7 YRVWFTSWSPCYD--CARHVDPLRGPNLSLIFARLYFCEDRAEP---EGLRLH 130  
QY 131 RAGVOIATMTFQDYFCNFTFVENHRTFKAMEGLHNSVRLSRQLRIL 180  
DB 62 DAGAQSIVTITDEFYICMDTFVYRQCGCFQPMWGLIEHSQALSGRLRAIL 111

RESULT 8  
US-09-128-395-3  
Sequence 3, Application US/09128395  
Patent No. 6087108  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Goli, Surya K.  
TITLE OF INVENTION: NOVEL RNA EDITING ENZYME  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/128,395  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/816,241  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0239 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 116 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 436941  
US-09-128-395-3

Query Match 22.6%; Score 245; DB 3; Length 116;  
Best Local Similarity 46.4%; Pred. No. 5.1e-21;  
Matches 51; Conservative 17; Mismatches 32; Indels 10; Gaps 3;

QY 76 YRVWFTSWSPCYD--CARHVDPLRGPNLSLIFARLYFCEDRAEP---EGLRLH 130  
DB 7 YRVWFTSWSPCYD--CARHVDPLRGPNLSLIFARLYFCEDRAEP---EGLRLH 130  
QY 131 RAGVOIATMTFQDYFCNFTFVENHRTFKAMEGLHNSVRLSRQLRIL 180  
DB 62 DAGAQSIVTITDEFYICMDTFVYRQCGCFQPMWGLIEHSQALSGRLRAIL 111

RESULT 9  
US-09-079-030-224



[illegible]



ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0239 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 236 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 11777906  
US-09-816-241-4

Query Match 19.9%; Score 216.5; DB 1; Length 236;  
Best Local Similarity 34.9%; Pred. No. 3e-17;  
Matches 44; Conservative 31; Mismatches 42; Indels 9; Gaps 4;

QY 35 RDSATSPSLDFGYLR-----NKGQC-HVELLFL-RYISPDWLDLPGRCYRTWTSP 86  
DB 33 REACCLLEIKMGMSKIRSSGKNTTNHVEVNFIKFTSERDFPISCSITWFLSWSP 92

QY 87 CYDCARHVADFLRGNNLSLRIFTARLYFCEDRKAEPGRLRLHRAQVQIAIMTFKDYFY 146  
DB 93 CMECSQAIREFLSRHGGTVLVYVARLFWHMDQ-NRQGLRLVNSGVTIQLMRASEYXH 151

QY 147 CWNTEY 152  
DB 152 CWNTEY 157

RESULT 12  
US-09-040-482-4  
Sequence 4, Application US/09040482  
Patent No. 5916556  
GENERAL INFORMATION:  
APPLICANT: Au-Young, Janice  
APPLICANT: Hawkins, Phillip R.  
APPLICANT: Hillman, Jennifer L.  
TITLE OF INVENTION: A NOVEL HUMAN MRNA EDITING ENZYME  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: U.S.  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/040,482  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/687,895  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0109 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 236 amino acids

TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 1177798  
US-09-040-482-4

Query Match 19.9%; Score 216.5; DB 2; Length 236;  
Best Local Similarity 34.9%; Pred. No. 3e-17;  
Matches 44; Conservative 31; Mismatches 42; Indels 9; Gaps 4;

QY 35 RDSATSPSLDFGYLR-----NKGQC-HVELLFL-RYISPDWLDLPGRCYRTWTSP 86  
DB 33 REACCLLEIKMGMSKIRSSGKNTTNHVEVNFIKFTSERDFPISCSITWFLSWSP 92

QY 87 CYDCARHVADFLRGNNLSLRIFTARLYFCEDRKAEPGRLRLHRAQVQIAIMTFKDYFY 146  
DB 93 CMECSQAIREFLSRHGGTVLVYVARLFWHMDQ-NRQGLRLVNSGVTIQLMRASEYXH 151

QY 147 CWNTEY 152  
DB 152 CWNTEY 157

RESULT 13  
US-09-128-395-4  
Sequence 4, Application US/09128395  
Patent No. 6087108  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Goll, Surya K.  
TITLE OF INVENTION: NOVEL RNA EDITING ENZYME  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/128,395  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/816,241  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0239 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 236 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 11777906  
US-09-128-395-4

Query Match 19.9%; Score 216.5; DB 3; Length 236;



Best Local Similarity 34.9%; Pred. No. 3e-17;  
Matches 44; Conservative 31; Mismatches 42; Indels 9; Gaps 4;

QY 35 RRDATSFSLDPGYLR-----NKNGC-HVELLFL-RYISDMDLDGRCYRTWFSWSP 86  
Db 33 RKEACLIYEIKWMSKRMKSSGKNTTNHVENFIIKFTSERDHPFSISCSITWFLMSNP 92  
QY 87 CYDCARHVAADFLRGPNLSIRIFTARLYFCEDRKAPBGLRLHRAVQIAIMTFKDYF 146  
Db 93 CWFCSQAIRFELSRHPVTLIVYARLFMMDDQ-NRQGLRDLYNSGVITQIMRASEYH 151  
QY 147 CWNTEFV 152  
Db 152 CWNTEFV 157

RESULT 14  
US-08-158-682A-4  
Sequence 4, Application US/08158682A  
Patent No. 5434058  
GENERAL INFORMATION:  
APPLICANT: Davidson, Nicholas O.  
TITLE OF INVENTION: Apolipoprotein B RNA Editing Protein:  
NUMBER OF INVENTIONS: Composition and Method  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ARNOLD, WHITE & DURKEE  
STREET: 321 No. 5434058th Clark Street, Suite 800  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60610  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/158,682A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Coolley, Ronald B.  
REGISTRATION NUMBER: 27,187  
REFERENCE/DOCKET NUMBER: ARCD:085  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 744-0090  
TELEFAX: (312) 245-4961  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 236 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-158-682A-4

Query Match 19.3%; Score 210; DB 1; Length 236;  
Best Local Similarity 35.6%; Pred. No. 1.7e-16;  
Matches 48; Conservative 28; Mismatches 43; Indels 16; Gaps 5;

QY 24 RRETYLCYVK-----RRDSATSFSLDPGYLRNKNGC-HVELLFL-RYISDMDLDGRCYR 77  
Db 33 RKEACLIYEIKWMSKRMKSSGKNTTNHVENFIIKFTSERDHPFSISCSITWFLMSNP 83  
QY 78 VTWFTSWSPCYDCARHVAADFLRGPNLSIRIFTARLYFCEDRKAPBGLRLHRAVQIA 137  
Db 84 ITWLSWSPCWFCSQAIRFELSRHPVTLIVYARLFMMDDQ-NRQGLRDLYNSGVITQ 142  
QY 138 IMTFKDYFCWNTEFV 152  
Db 143 IMRASEYHCKWNTEFV 157

RESULT 15  
US-08-158-682A-2  
Sequence 2, Application US/08158682A  
Patent No. 5434058  
GENERAL INFORMATION:  
APPLICANT: Davidson, Nicholas O.  
TITLE OF INVENTION: Apolipoprotein B RNA Editing Protein:  
NUMBER OF INVENTIONS: Composition and Method  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ARNOLD, WHITE & DURKEE  
STREET: 321 No. 5434058th Clark Street, Suite 800  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60610  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/158,682A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Coolley, Ronald B.  
REGISTRATION NUMBER: 27,187  
REFERENCE/DOCKET NUMBER: ARCD:085  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 744-0090  
TELEFAX: (312) 245-4961  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 229 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-158-682A-2

Query Match 18.2%; Score 198; DB 1; Length 229;  
Best Local Similarity 32.9%; Pred. No. 4.2e-15;  
Matches 56; Conservative 24; Mismatches 62; Indels 28; Gaps 6;

QY 24 RRETYLCYVK-----RRDSATSFSLDPGYLRNKNGCHVELLFL-RYISDMDLDGRCYR 74  
Db 33 RRETCLEIWMGSHSIWRHSQNT-----NK---HVENFIEKFTSERDHPFSISCSITWFLMSNP 80  
QY 75 CVRTWFTSWSPCYDCARHVAADFLRGPNLSIRIFTARLYFCEDRKAPBGLRLHRAVQIA 134  
Db 81 RCSIWFTSWSPCWFCSQAIRFELSRHPVTLIVYARLFMMDDQ-NRQGLRDLYNSGV 139  
QY 135 QIAIMTFKDYFCWNTEFVFNHRTFKAWEGHNSVRLSRQLRILPLY 184  
Db 140 TIQIMTEBSGVCWNTEFVNSPNSNAHWPRIYHMLVRL-----YVLELY 183

Search completed: March 11, 2004, 17:34:57  
Job time : 23 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Comogen Ltd.

OM protein - protein search, using sw model

Run on: March 11, 2004, 17:34:20 ; Search time 34 Seconds  
(without alignments)  
1229.657 Million cell updates/sec

Title: US-09-966-880A-8

Perfect score: 1086  
Sequence: 1 MSLLMKRRKFLYQKVRW.....ILLFLYVDLDAFRTGL 198

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 21153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:  
1: /cgn2\_6/ptodata/1/pubppa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubppa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubppa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubppa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubppa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubppa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubppa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubppa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubppa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubppa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubppa/US09\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/1/pubppa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/1/pubppa/US10\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubppa/US10\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/1/pubppa/US10\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubppa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/1/pubppa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/1/pubppa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1086	100.0	198	US-09-966-880A-8	Sequence 8, Appl
2	1008	92.8	198	US-09-966-880A-2	Sequence 2, Appl
3	390	35.9	189	US-10-460-923-5	Sequence 5, Appl
4	390	35.9	384	US-09-729-674-174	Sequence 174, App
5	390	35.9	384	US-10-460-923-2	Sequence 2, Appl
6	369.5	34.0	222	US-09-925-300-1639	Sequence 1639, Ap
7	363.5	33.5	199	US-10-460-923-7	Sequence 7, Appl
8	349	32.1	210	US-10-460-923-4	Sequence 4, Appl
9	300	27.6	152	US-10-247-671-159	Sequence 159, App
10	238.5	22.0	195	US-10-460-923-3	Sequence 3, Appl
11	230	21.2	219	US-10-460-923-6	Sequence 6, Appl
12	218.5	20.1	236	US-10-157-031-14	Sequence 14, Appl
13	216.5	19.9	236	US-10-460-923-8	Sequence 8, Appl
14	211	19.4	229	US-09-966-880A-36	Sequence 36, Appl
15	198	18.2	127	US-10-104-047-3729	Sequence 3729, Ap

16	167	15.4	128	15	US-10-378-029-77	Sequence 77, Appl
17	154	14.2	151	14	US-10-029-386-34155	Sequence 34155, A
18	84	7.7	51	9	US-09-864-761-38853	Sequence 38853, A
19	81	7.5	440	13	US-10-120-319-3	Sequence 3, Appl
20	81	7.5	440	14	US-10-189-977-3	Sequence 3, Appl
21	81	7.5	440	14	US-10-392-098-3	Sequence 3, Appl
22	80	7.4	476	9	US-09-733-524-15	Sequence 15, Appl
23	80	7.4	476	13	US-10-120-319-5	Sequence 5, Appl
24	80	7.4	476	14	US-10-189-977-5	Sequence 5, Appl
25	80	7.4	476	14	US-10-392-098-5	Sequence 5, Appl
26	77.5	7.1	328	15	US-09-851-873-55	Sequence 55, Appl
27	77.5	7.1	328	15	US-10-369-493-6748	Sequence 6748, Ap
28	77.5	7.1	663	13	US-10-080-960-14	Sequence 14, Appl
29	77.5	7.1	663	14	US-10-247-671-135	Sequence 135, App
30	77	7.1	790	14	US-10-153-668-1164	Sequence 164, App
31	75.5	7.0	382	10	US-09-947-208-25	Sequence 25, Appl
32	75	6.9	330	14	US-10-265-593-4	Sequence 4, Appl
33	74	6.8	200	10	US-09-851-873-65	Sequence 65, Appl
34	74	6.8	707	15	US-10-014-099F-61	Sequence 61, Appl
35	74	6.8	1291	15	US-10-452-024-122	Sequence 122, App
36	73	6.7	361	15	US-10-439-741-2	Sequence 2, Appl
37	72	6.6	311	16	US-10-354-437-4	Sequence 4, Appl
38	72	6.6	357	14	US-10-285-976-35	Sequence 35, Appl
39	72	6.6	361	15	US-10-312-088-27	Sequence 27, Appl
40	72	6.6	363	14	US-10-038-288A-4	Sequence 4, Appl
41	72	6.6	433	14	US-10-038-288A-2	Sequence 2, Appl
42	72	6.6	704	15	US-10-312-088-26	Sequence 26, Appl
43	72	6.6	1291	15	US-10-452-024-121	Sequence 121, App
44	72	6.6	1291	15	US-10-452-024-123	Sequence 123, App
45	71.5	6.6	529	14	US-10-053-192-5	Sequence 5, Appl

## ALIGNMENTS

RESULT 1  
US-09-966-880A-8  
; Sequence 8, Application US/09966880A  
; Patent No. US20020164743A1  
; GENERAL INFORMATION:  
; APPLICANT: Honjo, Yasuko  
; APPLICANT: Muramatsu, Masamichi  
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
; FILE REFERENCE: 06501-088001  
; CURRENT APPLICATION NUMBER: US/09/966, 880A  
; CURRENT FILING DATE: 2001-09-28  
; PRIOR APPLICATION NUMBER: PCT/JP00/01918  
; PRIOR FILING DATE: 2000-03-28  
; PRIOR APPLICATION NUMBER: JP 11-371382  
; PRIOR FILING DATE: 1999-12-27  
; PRIOR APPLICATION NUMBER: JP 11-178999  
; PRIOR FILING DATE: 1999-06-24  
; PRIOR APPLICATION NUMBER: JP 11-87192  
; PRIOR FILING DATE: 1999-03-29  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8  
; LENGTH: 198  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-966-880A-8

Query Match 100.0%; Score 1086; DB 9; Length 198;  
Best Local Similarity 100.0%; Pred. No. 6.1e-114;  
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MSLLMKRRKFLYQKVRWAKGRRETLGVVRRBDSATSFSDGFLRKNGCHVELL	60
DB	1	MSLLMKRRKFLYQKVRWAKGRRETLGVVRRBDSATSFSDGFLRKNGCHVELL	60
QY	61	FLRYISDMDLPGRRCYRTWTFTSWSPCYCARHVAADFLRGNPNLSLRTFLRLYFCEDRK	120
DB	61	FLRYISDMDLPGRRCYRTWTFTSWSPCYCARHVAADFLRGNPNLSLRTFLRLYFCEDRK	120



QY 121 AEEGRLRLHRAAGVQAIAMTFKDYFCWNTFVENHRTFKAWEGHLSVRLSRQLRIL 180  
DB 121 AEEGRLRLHRAAGVQAIAMTFKDYFCWNTFVENHRTFKAWEGHLSVRLSRQLRIL 180  
QY 181 LPLVEVDLRLDAFRTGL 198  
DB 181 LPLVEVDLRLDAFRTGL 198

## RESULT 2

US-09-966-880A-2  
Sequence 2, Application US/09966880A  
Patent No. US20020164743A1  
GENERAL INFORMATION:  
APPLICANT: Honjo, Tasuku  
APPLICANT: Muramatsu, Masamichi  
TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
FILE REFERENCE: 06501-088001  
CURRENT FILING DATE: 2001-09-28  
PRIOR APPLICATION NUMBER: PCT/JP00/01918  
PRIOR FILING DATE: 2000-03-28  
PRIOR APPLICATION NUMBER: JP 11-371382  
PRIOR FILING DATE: 1999-12-27  
PRIOR APPLICATION NUMBER: JP 11-178999  
PRIOR FILING DATE: 1999-06-24  
PRIOR APPLICATION NUMBER: JP 11-87192  
PRIOR FILING DATE: 1999-03-29  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 2  
LENGTH: 198  
TYPE: PRT  
ORGANISM: Mus musculus  
US-09-966-880A-2

Query Match 92.8%; Score 1008; DB 9; Length 198;  
Best Local Similarity 92.9%; Pred. No. 3,66-105;  
Matches 183; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 1 MDSLIMNRKRLKLYQKQVNRWAKGRRETYLCYVVRKRSATSFSLD-FGYLRNKNGCHVELL 60  
DB 1 MDSLIMNRKRLKLYQKQVNRWAKGRRETYLCYVVRKRSATSFSLD-FGYLRNKNGCHVELL 60  
QY 61 FLRYISDWDLDGRCYRTWTFTSWSPCYDCARHVAEFLRNPNLSLIFTRKLYFCEDRK 120  
DB 61 FLRYISDWDLDGRCYRTWTFTSWSPCYDCARHVAEFLRNPNLSLIFTRKLYFCEDRK 120  
QY 121 AEEGRLRLHRAAGVQAIAMTFKDYFCWNTFVENHRTFKAWEGHLSVRLSRQLRIL 180  
DB 121 AEEGRLRLHRAAGVQAIAMTFKDYFCWNTFVENHRTFKAWEGHLSVRLSRQLRIL 180  
QY 181 LPLVEVDLRLDAFRTGL 197  
DB 181 LPLVEVDLRLDAFRTGL 197

## RESULT 3

US-10-460-923-5  
Sequence 5, Application US/10460923  
Publication No. US2004000951A1  
GENERAL INFORMATION:  
APPLICANT: MALIM, Michael H.  
APPLICANT: SHEEHY, Ann M.  
APPLICANT: HARRIS, Reuben S.  
APPLICANT: BISHOP, Kate N.  
APPLICANT: NEUBERGER, Michael S.  
APPLICANT: GADDIS, Nathan C.  
APPLICANT: SIMON, James H.M.  
TITLE OF INVENTION: DNA Dendrimation Mediates Innate Immunity to Retroviral Infection  
FILE REFERENCE: 22253-74380  
CURRENT APPLICATION NUMBER: US/10/460,923

QY 6 MNRKRLKLYQKQVNRWAKGRRETYLCYVVRKRSATSFSLD-FGYLRNKNGCHVELL 55  
DB 2 MDPFTFTFNPNNEPVRGRHETLYCYEVERHNDTWLILNORRGFLCNQAPHKHGFLEGR 61  
QY 56 HVELLFLRYISDWDLDGRCYRTWTFTSWSPCYDCARHVAEFLRNPNLSLIFTRKLYFCEDRK 115  
DB 62 HVELLFLRYISDWDLDGRCYRTWTFTSWSPCYDCARHVAEFLRNPNLSLIFTRKLYFCEDRK 120  
QY 116 CEDRAEPEGRLRLHRAAGVQAIAMTFKDYFCWNTFVENHRTFKAWEGHLSVRLSRQLRIL 175  
DB 121 -DDGRCCEGLRTLAAGAKISIMTYSEFKACMDTFDHDGCPQPMDDGHDHSDLSGR 179  
QY 176 LRLIL 180  
DB 180 LRLIL 184

Query Match 35.9%; Score 390; DB 15; Length 189;  
Best Local Similarity 44.9%; Pred. No. 1,28-35;

Matches 83; Conservative 31; Mismatches 59; Indels 12; Gaps 3;

## RESULT 4

US-09-729-674-174  
Sequence 174, Application US/09729674  
Patent No. US20010039335A1  
GENERAL INFORMATION:  
APPLICANT: Jacobs, Kenneth  
APPLICANT: McCoy, John M.  
APPLICANT: Lavallee, Edward R.  
APPLICANT: Collins-Racie, Lisa A.  
APPLICANT: Evans, Cheryl  
APPLICANT: Weidberg, David  
APPLICANT: Treacy, Maurice  
APPLICANT: Agostino, Michael J.  
APPLICANT: Steindinger II, Robert J.  
APPLICANT: Spaulding, Vikki  
APPLICANT: Wong, Gordon G.  
APPLICANT: Clark, Hilary  
APPLICANT: Fecthel, Kim  
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM  
FILE REFERENCE: 6055-64X  
CURRENT FILING DATE: 2000-12-04  
PRIOR APPLICATION NUMBER: 09/539,330  
PRIOR FILING DATE: 2000-03-30  
NUMBER OF SEQ ID NOS: 263  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 174  
LENGTH: 384  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-729-674-174

Query Match 35.9%; Score 390; DB 9; Length 384;  
Best Local Similarity 44.9%; Pred. No. 2,88-35;  
Matches 83; Conservative 31; Mismatches 59; Indels 12; Gaps 3;

QY 6 MNRKRLKLYQKQVNRWAKGRRETYLCYVVRKRSATSFSLD-FGYLRNKNGCHVELL 55  
DB 121 -DDGRCCEGLRTLAAGAKISIMTYSEFKACMDTFDHDGCPQPMDDGHDHSDLSGR 179



```
Db 197 MDPTFFNFNNPFWRGHEHYLCYEVZRMKNDTWLLNQRGLCNQAPRHGFLGCR 256
QY 56 HVELLFRLYISDMDLPGRCYRTWFTSWSPCYDCARHVADEFARGNPNSLRIFARLYF 115
Db 257 HAEELCFDLVTPFWKLDLDDYRTCTFSWSPCSCQAEAKISKNKXVSLCIFARLY- 315
QY 116 CEDRKAPEGLRLHRAVQOIALMTFKDYFCWNTFVENHERTFKAMEGLHENSVALSRQ 175
Db 316 -DDGRCQEGRLTLAAGAKISIMTYSEFKHCWDTFVDHGGCPFPQMDLDSHSDLSGR 374
QY 176 LRRIL 180
Db 375 LRAIL 379

RESULT 5
US-10-460-923-2
; Sequence 2, Application US/10460923
; Publication No. US20040009951A1
; GENERAL INFORMATION:
; APPLICANT: MALIM, Michael H.
; APPLICANT: SHEEHY, Ann M.
; APPLICANT: HARRIS, Reuben S.
; APPLICANT: BISHOP, Kate N.
; APPLICANT: NEUBERGER, Michael S.
; APPLICANT: GADDIS, Nathan C.
; APPLICANT: SIMON, James H.M.
; TITLE OF INVENTION: DNA Deamination Mediates Innate Immunity to Retroviral Infection
; FILE REFERENCE: 22253-74380
; CURRENT APPLICATION NUMBER: US/10/460,923
; CURRENT FILING DATE: 2003-06-13
; PRIOR APPLICATION NUMBER: US 60/388,513
; PRIOR FILING DATE: 2002-06-13
; PRIOR APPLICATION NUMBER: US 60/472,952
; PRIOR FILING DATE: 2003-05-23
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 384
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-460-923-2

Query Match 35.9%; Score 390; DB 15; Length 384;
Best Local Similarity 44.9%; Pred. No. 2.8e-35;
Matches 83; Conservative 31; Mismatches 59; Indels 12; Gaps 3;

QY 6 MNRKFLYQFKVNAWMAKGRRETYLCYVVKRDSATSFSLD--FGYLNRK-----NCG 55
Db 197 MDPTFFNFNNPFWRGHEHYLCYEVZRMKNDTWLLNQRGLCNQAPRHGFLGCR 256
QY 56 HVELLFRLYISDMDLPGRCYRTWFTSWSPCYDCARHVADEFARGNPNSLRIFARLYF 115
Db 257 HAEELCFDLVTPFWKLDLDDYRTCTFSWSPCSCQAEAKISKNKXVSLCIFARLY- 315
QY 116 CEDRKAPEGLRLHRAVQOIALMTFKDYFCWNTFVENHERTFKAMEGLHENSVALSRQ 175
Db 316 -DDGRCQEGRLTLAAGAKISIMTYSEFKHCWDTFVDHGGCPFPQMDLDSHSDLSGR 374
QY 176 LRRIL 180
Db 375 LRAIL 379

RESULT 6
US-09-925-300-1639
; Sequence 1639, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruden,
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
```

```
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1639
; LENGTH: 222
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-300-1639

Query Match 34.0%; Score 369.5; DB 9; Length 222;
Best Local Similarity 44.9%; Pred. No. 2.9e-33;
Matches 79; Conservative 24; Mismatches 64; Indels 9; Gaps 4;

QY 11 FLYQFKVNAWMAKGRRETYLCYVVK--RDSATSFSLDGYLRN---KNGCHVELLFLRY 64
Db 49 FYQFKVNAWMAKGRRETYLCYVVK--RDSATSFSLDGYLRN---KNGCHVELLFLRY 106
QY 65 ISDMDLPGRCYRTWFTSWSPCYDCARHVADEFARGNPNSLRIFARLYFCEBRKAPE 124
Db 107 FCDLILSPNTKYQYTWFTSWSPCYDCARHVADEFARGNPNSLRIFARLYFQ--YPCYOE 165
QY 125 GLRLHRAVQOIALMTFKDYFCWNTFVENHERTFKAMEGLHENSVALSRQLRIL 180
Db 166 GLRLHRAVQOIALMTFKDYFCWNTFVENHERTFKAMEGLHENSVALSRQLRIL 221

RESULT 7
US-10-460-923-7
; Sequence 7, Application US/10460923
; Publication No. US20040009951A1
; GENERAL INFORMATION:
; APPLICANT: MALIM, Michael H.
; APPLICANT: SHEEHY, Ann M.
; APPLICANT: HARRIS, Reuben S.
; APPLICANT: BISHOP, Kate N.
; APPLICANT: NEUBERGER, Michael S.
; APPLICANT: GADDIS, Nathan C.
; APPLICANT: SIMON, James H.M.
; TITLE OF INVENTION: DNA Deamination Mediates Innate Immunity to Retroviral Infection
; FILE REFERENCE: 22253-74380
; CURRENT APPLICATION NUMBER: US/10/460,923
; CURRENT FILING DATE: 2003-06-13
; PRIOR APPLICATION NUMBER: US 60/388,513
; PRIOR FILING DATE: 2002-06-13
; PRIOR APPLICATION NUMBER: US 60/472,952
; PRIOR FILING DATE: 2003-05-23
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 199
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-460-923-7

Query Match 33.5%; Score 363.5; DB 15; Length 199;
Best Local Similarity 43.5%; Pred. No. 1.2e-32;
Matches 83; Conservative 28; Mismatches 57; Indels 23; Gaps 7;

QY 5 LMRKFLYQFKVNAWMAKGRRETYLCYVVKRDSATSFSLD--FGYLNRK-----G 54
Db 12 LMDPHILFNSFRN---GGRKTYLCYEVZRMKNDTWLLNQRGLCNQAPRHGFLGCR 68
QY 55 HVELLFRLYISDMDLPGRCYRTWFTSWSPCYDCARHVADEFARGNPNSLRIFARLYF 112
Db 69 HAEELCFDLVTPFWKLDLDDYRTCTFSWSPCSCQAEAKISKNKXVSLCIFARLY- 128
QY 113 LYFCEDRAEP---EGLRLHRAVQOIALMTFKDYFCWNTFVENHERTFKAMEGLHENS 169
```



Db 129 IY-----DYDPLYKEALQWLRDAGAVSINTYDEFKHCMDTFYDHQGCFFQFWDGLDHS 183  
QY 170 VRLSRQLRLL 180  
Db 184 QALSGRLRAIL 194

RESULT 8  
US-10-460-923-4  
Sequence 4, Application US/10460923  
Publication No. US2004000951A1  
GENERAL INFORMATION:  
APPLICANT: MALIN, Michael H.  
APPLICANT: SHEEHY, Ann M.  
APPLICANT: HARRIS, Reuben S.  
APPLICANT: BISHOP, Kate N.  
APPLICANT: NEUBERGER, Michael S.  
APPLICANT: GADDIS, Nathan C.  
APPLICANT: SIMON, James H.M.  
TITLE OF INVENTION: DNA Deamination Mediates Innate Immunity to Retroviral Infection  
FILE REFERENCE: 22253-74380  
CURRENT APPLICATION NUMBER: US/10/460,923  
CURRENT FILING DATE: 2003-06-13  
PRIOR APPLICATION NUMBER: US 60/388,513  
PRIOR FILING DATE: 2002-06-13  
PRIOR APPLICATION NUMBER: US 60/472,952  
PRIOR FILING DATE: 2003-05-23  
NUMBER OF SEQ ID NOS: 12  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 4  
LENGTH: 210  
TYPE: PRT  
ORGANISM: Unknown  
FEATURE:  
OTHER INFORMATION: mouse orthologue  
US-10-460-923-4

Query Match 32.1%; Score 349; DB 15; Length 210;  
Best Local Similarity 38.8%; Pred. No. 5.4e-31;  
Matches 71; Conservative 36; Mismatches 72; Indels 4; Gaps 2;

QY 5 LMRKKFLYQFKNVMAKGRRTYLCYVVKRDSATSSLDPGYIRNKGCHVLLFRY 64  
Db 20 LISQETFKFHFRLRYALDRKDTFLCYEYTRKDCSPVSLHGVKXNDTHAICFLYW 79  
QY 65 ISD---WLDPRGRCRYVMTSMPSCYDCARHVADFLGNPNLSLRIFARLYFCEDRKA 121  
Db 80 FHDKYLKVLSPREERKLTWMSWPCFBCAGVLFLLATHNLSLDITSSRLYNIRDFEN 139  
QY 122 EPEGRRLHRAVQVIAIMTFKDYFCWNTFVENHRTKAMEGLHNSVRLSRQLRLL 181  
Db 140 Q-QNLRLVQGAQVAMDLVEFKCKWKFFVNGRRPRPMKLLTNFRYQDSKLQELLR 198

QY 182 PLY 184  
Db 199 PCY 201

RESULT 9  
US-10-247-671-159  
Sequence 159, Application US/10247671  
Publication No. US20030194721A1  
GENERAL INFORMATION:  
APPLICANT: Wikita, Thomas  
APPLICANT: Shiffman, Dov  
APPLICANT: Porter, Gordon, J.  
APPLICANT: Kaser, Matthew R.  
TITLE OF INVENTION: GENES EXPRESSED IN TREATED FOAM CELLS  
FILE REFERENCE: PA-0050 US  
CURRENT APPLICATION NUMBER: US/10/247,671  
CURRENT FILING DATE: 2002-09-18  
PRIOR APPLICATION NUMBER: 60/323,784  
PRIOR FILING DATE: 2001-09-19

NUMBER OF SEQ ID NOS: 186  
SOFTWARE: PERL Program  
SEQ ID NO 159  
LENGTH: 152  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: Incyte ID No. US20030194721A1 135626CD1  
US-10-247-671-159

Query Match 27.6%; Score 300; DB 14; Length 152;  
Best Local Similarity 44.9%; Pred. No. 1.2e-25;  
Matches 66; Conservative 20; Mismatches 43; Indels 18; Gaps 5;

QY 47 GYLRN--KN-----GCHVEFLRLYISDWLDPGRCYRVTFWTSWPCYD--CARHVAD 96  
Db 6 GFLHQAKNLGCFYGRHAEIRFLDVLPSLDLDDRAQIRYRVTFWTSWPCFSGMGAQGVRA 65  
QY 97 PLRGNPNLSLRIFARLYFCEDRKAEP---EGLRLHRAVQVIAIMTFKDYFCWNTFVE 153  
Db 66 FLQENTHRLRLIFARLY-----DYDPLYKEALQWLRDAGAVSINTYDEFKHCMDTFVY 120  
QY 154 NHERTFKAMEGLHNSVRLSRQLRLL 180  
Db 121 RQGCFFQFWDGLHEHSQLSGRLRAIL 147

RESULT 10  
US-10-460-923-3  
Sequence 3, Application US/10460923  
Publication No. US2004000951A1  
GENERAL INFORMATION:  
APPLICANT: MALIN, Michael H.  
APPLICANT: SHEEHY, Ann M.  
APPLICANT: HARRIS, Reuben S.  
APPLICANT: BISHOP, Kate N.  
APPLICANT: NEUBERGER, Michael S.  
APPLICANT: GADDIS, Nathan C.  
APPLICANT: SIMON, James H.M.  
TITLE OF INVENTION: DNA Deamination Mediates Innate Immunity to Retroviral Infection  
FILE REFERENCE: 22253-74380  
CURRENT APPLICATION NUMBER: US/10/460,923  
CURRENT FILING DATE: 2003-06-13  
PRIOR APPLICATION NUMBER: US 60/388,513  
PRIOR FILING DATE: 2002-06-13  
PRIOR APPLICATION NUMBER: US 60/472,952  
PRIOR FILING DATE: 2003-05-23  
NUMBER OF SEQ ID NOS: 12  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 3  
LENGTH: 195  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-460-923-3

Query Match 22.0%; Score 238.5; DB 15; Length 195;  
Best Local Similarity 34.2%; Pred. No. 1.3e-18;  
Matches 63; Conservative 26; Mismatches 84; Indels 11; Gaps 6;

QY 6 MMRKKFLYQFKNVMAKGRRTYLCYVVKRDSATSSLDPGYIRNKGCHVLLFRY 61  
Db 12 MYRDTFSYNPNPILSRRTVWLCEVTKGSPRP-PLDAKTRQGVSELKTHPEMRF 70  
QY 62 LRYISDW-DLDPGRCYRVTFWTSWPCYDCARHVADFLGNPNLSLRIFARLYFCEDRKA 120  
Db 71 FHWFSKRRKLRHROEYVTVTYSMPSCYDCARHVADFLGNPNLSLRIFARLYFCEDRKA 130  
QY 121 APEGRRL--HRAV--QVIAIMTFKDYFCWNTFVENHRTKAMEGLHNSVRLSRQL 176  
Db 131 YQ-EALNSLQKQKDGPRATWKIMNYDBFQGCWKFVYSQELPEPMNLPKYYILLHIML 189  
QY 177 RRL 180



Db 190 GEIL 193

RESULT 11

US-10-460-923-6

Sequence 6, Application US/10460923

Publication No. US20040009951A1

GENERAL INFORMATION:

APPLICANT: MALIM, Michael H.

APPLICANT: SHEEHY, Ann M.

APPLICANT: HARRIS, Reuben S.

APPLICANT: BISHOP, Kate N.

APPLICANT: NEUBERGER, Michael S.

APPLICANT: GADDIS, Nathan C.

APPLICANT: SIMON, James H.M.

TITLE OF INVENTION: DNA Deamination Mediates Innate Immunity to Retroviral Infection

FILE REFERENCE: 22253-74380

CURRENT APPLICATION NUMBER: US/10/460,923

PRIOR FILING DATE: 2003-06-13

PRIOR APPLICATION NUMBER: US 60/388,513

PRIOR FILING DATE: 2002-06-13

PRIOR APPLICATION NUMBER: US 60/472,952

NUMBER OF SEQ ID NOS: 12

SOFTWARE: PatentIn version 3.1

SEQ ID NO 6

LENGTH: 219

TYPE: PRT

ORGANISM: Unknown

FEATURE:

OTHER INFORMATION: mouse orthologue

US-10-460-923-6

Query Match

Best Local Similarity 21.2%; Score 230; DB 15; Length 219;

Best Local Similarity 33.5%; Pred. No. 1.4e-17;

Matches 67; Conservative 35; Mismatches 82; Indels 16; Gaps 8;

Query Match

Best Local Similarity 21.2%; Score 230; DB 15; Length 219;

Best Local Similarity 33.5%; Pred. No. 1.4e-17;

Matches 67; Conservative 35; Mismatches 82; Indels 16; Gaps 8;

Query Match

Best Local Similarity 21.2%; Score 230; DB 15; Length 219;

Best Local Similarity 33.5%; Pred. No. 1.4e-17;

Matches 67; Conservative 35; Mismatches 82; Indels 16; Gaps 8;

Query Match

Best Local Similarity 21.2%; Score 230; DB 15; Length 219;

Best Local Similarity 33.5%; Pred. No. 1.4e-17;

Matches 67; Conservative 35; Mismatches 82; Indels 16; Gaps 8;

Query Match

Best Local Similarity 21.2%; Score 230; DB 15; Length 219;

Best Local Similarity 33.5%; Pred. No. 1.4e-17;

Matches 67; Conservative 35; Mismatches 82; Indels 16; Gaps 8;

Query Match

Best Local Similarity 21.2%; Score 230; DB 15; Length 219;

Best Local Similarity 33.5%; Pred. No. 1.4e-17;

Matches 67; Conservative 35; Mismatches 82; Indels 16; Gaps 8;

Query Match

Best Local Similarity 21.2%; Score 230; DB 15; Length 219;

Best Local Similarity 33.5%; Pred. No. 1.4e-17;

Matches 67; Conservative 35; Mismatches 82; Indels 16; Gaps 8;

Query Match

Best Local Similarity 21.2%; Score 230; DB 15; Length 219;

Best Local Similarity 33.5%; Pred. No. 1.4e-17;

Matches 67; Conservative 35; Mismatches 82; Indels 16; Gaps 8;

Query Match

Best Local Similarity 21.2%; Score 230; DB 15; Length 219;

Best Local Similarity 33.5%; Pred. No. 1.4e-17;

Matches 67; Conservative 35; Mismatches 82; Indels 16; Gaps 8;

Query Match

Best Local Similarity 21.2%; Score 230; DB 15; Length 219;

Best Local Similarity 33.5%; Pred. No. 1.4e-17;

Matches 67; Conservative 35; Mismatches 82; Indels 16; Gaps 8;

Query Match

Best Local Similarity 21.2%; Score 230; DB 15; Length 219;

Best Local Similarity 33.5%; Pred. No. 1.4e-17;

Matches 67; Conservative 35; Mismatches 82; Indels 16; Gaps 8;

Query Match

Best Local Similarity 21.2%; Score 230; DB 15; Length 219;

TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-157-031-14

Query Match

Best Local Similarity 20.1%; Score 218.5; DB 14; Length 236;

Best Local Similarity 34.9%; Pred. No. 3e-16;

Matches 44; Conservative 31; Mismatches 42; Indels 9; Gaps 4;

Query Match

Best Local Similarity 20.1%; Score 218.5; DB 14; Length 236;

Best Local Similarity 34.9%; Pred. No. 3e-16;

Matches 44; Conservative 31; Mismatches 42; Indels 9; Gaps 4;

Query Match

Best Local Similarity 20.1%; Score 218.5; DB 14; Length 236;

Best Local Similarity 34.9%; Pred. No. 3e-16;

Matches 44; Conservative 31; Mismatches 42; Indels 9; Gaps 4;

Query Match

Best Local Similarity 20.1%; Score 218.5; DB 14; Length 236;

Best Local Similarity 34.9%; Pred. No. 3e-16;

Matches 44; Conservative 31; Mismatches 42; Indels 9; Gaps 4;

Query Match

Best Local Similarity 20.1%; Score 218.5; DB 14; Length 236;

Best Local Similarity 34.9%; Pred. No. 3e-16;

Matches 44; Conservative 31; Mismatches 42; Indels 9; Gaps 4;

Query Match

Best Local Similarity 20.1%; Score 218.5; DB 14; Length 236;

Best Local Similarity 34.9%; Pred. No. 3e-16;

Matches 44; Conservative 31; Mismatches 42; Indels 9; Gaps 4;

Query Match

Best Local Similarity 20.1%; Score 218.5; DB 14; Length 236;

Best Local Similarity 34.9%; Pred. No. 3e-16;

Matches 44; Conservative 31; Mismatches 42; Indels 9; Gaps 4;

Query Match

Best Local Similarity 20.1%; Score 218.5; DB 14; Length 236;

Best Local Similarity 34.9%; Pred. No. 3e-16;

Matches 44; Conservative 31; Mismatches 42; Indels 9; Gaps 4;

Query Match

Best Local Similarity 20.1%; Score 218.5; DB 14; Length 236;

Best Local Similarity 34.9%; Pred. No. 3e-16;

Matches 44; Conservative 31; Mismatches 42; Indels 9; Gaps 4;

Query Match

Best Local Similarity 20.1%; Score 218.5; DB 14; Length 236;

Best Local Similarity 34.9%; Pred. No. 3e-16;

Matches 44; Conservative 31; Mismatches 42; Indels 9; Gaps 4;

Query Match

Best Local Similarity 20.1%; Score 218.5; DB 14; Length 236;

Best Local Similarity 34.9%; Pred. No. 3e-16;

Matches 44; Conservative 31; Mismatches 42; Indels 9; Gaps 4;

Query Match

Best Local Similarity 20.1%; Score 218.5; DB 14; Length 236;

Best Local Similarity 34.9%; Pred. No. 3e-16;

Matches 44; Conservative 31; Mismatches 42; Indels 9; Gaps 4;

Query Match

Best Local Similarity 20.1%; Score 218.5; DB 14; Length 236;

Best Local Similarity 34.9%; Pred. No. 3e-16;

Matches 44; Conservative 31; Mismatches 42; Indels 9; Gaps 4;

Query Match

Best Local Similarity 20.1%; Score 218.5; DB 14; Length 236;

Best Local Similarity 34.9%; Pred. No. 3e-16;

Matches 44; Conservative 31; Mismatches 42; Indels 9; Gaps 4;

Query Match

Best Local Similarity 20.1%; Score 218.5; DB 14; Length 236;

Best Local Similarity 34.9%; Pred. No. 3e-16;

Matches 44; Conservative 31; Mismatches 42; Indels 9; Gaps 4;

Query Match

Best Local Similarity 20.1%; Score 218.5; DB 14; Length 236;

Best Local Similarity 34.9%; Pred. No. 3e-16;

Matches 44; Conservative 31; Mismatches 42; Indels 9; Gaps 4;

Query Match

Best Local Similarity 20.1%; Score 218.5; DB 14; Length 236;

Best Local Similarity 34.9%; Pred. No. 3e-16;

Matches 44; Conservative 31; Mismatches 42; Indels 9; Gaps 4;

Query Match

Best Local Similarity 20.1%; Score 218.5; DB 14; Length 236;

Best Local Similarity 34.9%; Pred. No. 3e-16;

Matches 44; Conservative 31; Mismatches 42; Indels 9; Gaps 4;

Query Match

Best Local Similarity 20.1%; Score 218.5; DB 14; Length 236;

Best Local Similarity 34.9%; Pred. No. 3e-16;

Matches 44; Conservative 31; Mismatches 42; Indels 9; Gaps 4;







GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: March 13, 2004, 09:16:37 ; Search time 66 Seconds  
(without alignments)  
1664.853 Million cell updates/sec

Title: US-09-966-880A-8  
Perfect score: 1086  
Sequence: 1 MDSLLMNRKRLYQFKVNRW.....ILPLVEVDLDRDAFRTGL 198

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues  
Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+ p2n.model -DEV=xlh  
-O=/cg2\_1/USPTO.spool/US09966880/runat\_11032004\_172742\_26608/app\_query.fasta\_1.391  
-DB=Issued\_Patents\_NA -QPWT=fastap -SUFFIX=rnt -MINMATCH=0.1 -LOOPEC=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdt  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09966880 @CGN 1.1 44 @runat\_11032004\_172742\_26608 -NCU=6 -ICPU=3  
-NO WMAP -LARGEOUTERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGIOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:\*  
1: /cg2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cg2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cg2\_6/ptodata/2/ina/6A.COMB.seq:\*  
4: /cg2\_6/ptodata/2/ina/6B.COMB.seq:\*  
5: /cg2\_6/ptodata/2/ina/6C.COMB.seq:\*  
6: /cg2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	369.5	34.0	610	1	US-08-816-241-2
2	369.5	34.0	610	3	US-09-128-395-2
3	271	25.0	891	1	US-08-687-895-2
4	271	25.0	891	2	US-09-040-482-2
5	210	19.3	879	1	US-08-158-682A-3
6	201	18.5	879	1	US-08-158-682A-1
7	201	18.5	879	1	US-08-015-203-1
8	87.5	8.1	1806	3	US-09-068-655-10
9	87.5	8.1	4713	3	US-09-068-655-3
10	86	7.9	1230025	4	US-09-198-452A-1
11	80.5	7.4	1356	4	US-09-107-532A-2790
12	80	7.4	7077	4	US-09-221-017B-123

C	13	79.5	7.3	591	4	US-09-252-991A-7743	Sequence 7743, Ap
C	14	79.5	7.3	816	4	US-09-252-991A-7748	Sequence 7748, Ap
C	15	78	7.2	1053	4	US-09-252-991A-6143	Sequence 6143, Ap
C	16	78	7.2	1167	4	US-09-252-991A-6408	Sequence 6408, Ap
C	17	77.5	7.1	1209	6	5352575-4	Patent No. 5352575
C	18	77.5	7.1	1446	4	US-09-976-594-668	Sequence 668, App
C	19	76.5	7.0	1905	4	US-09-252-991A-7829	Sequence 7829, Ap
C	20	75.5	7.0	3043	4	US-09-049-698-16	Sequence 16, Appl
C	21	75.5	7.0	3181	4	US-09-049-698-18	Sequence 18, Appl
C	22	75	6.9	323	4	US-09-210-952-16	Sequence 16, Appl
C	23	74	6.8	269223	4	US-09-596-002-41	Sequence 41, Appl
C	24	73.5	6.8	852	4	US-09-488-035A-467	Sequence 467, App
C	25	73.5	6.8	2629	1	US-08-200-807-1	Sequence 1, Appl1
C	26	73.5	6.8	2629	1	US-08-488-305A-1	Sequence 1, Appl1
C	27	73	6.7	801	4	US-09-252-991A-2076	Sequence 2076, Ap
C	28	73	6.7	909	4	US-09-252-991A-2147	Sequence 2147, Ap
C	29	72.5	6.7	1578	4	US-09-107-532A-2089	Sequence 2089, Ap
C	30	72.5	6.7	9179	4	US-09-453-702B-162	Sequence 162, App
C	31	72.5	6.7	13188	4	US-08-961-527-70	Sequence 70, Appl
C	32	72	6.6	531	4	US-09-252-991A-4510	Sequence 4510, Ap
C	33	72	6.6	1464	4	US-09-489-035A-2639	Sequence 2639, Ap
C	34	72	6.6	1672	4	US-09-495-065-1	Sequence 1, Appl1
C	35	72	6.6	4201	1	US-08-080-255-4	Sequence 4, Appl1
C	36	72	6.6	4201	3	US-08-465-713-4	Sequence 4, Appl1
C	37	72	6.6	4201	5	PCT-US93-05857-4	Sequence 4, Appl1
C	38	72	6.6	5751	4	US-09-023-655-1415	Sequence 1415, Ap
C	39	72	6.6	11907	3	US-08-061-376-4	Sequence 4, Appl1
C	40	72	6.6	14255	1	US-08-320-559-1	Sequence 1, Appl1
C	41	72	6.6	14255	1	US-08-327-392-1	Sequence 1, Appl1
C	42	72	6.6	14255	1	US-08-306-691B-55	Sequence 55, Appl1
C	43	72	6.6	14255	3	US-08-545-860D-1	Sequence 1, Appl1
C	44	72	6.6	14255	5	PCT-US84-0446-1	Sequence 1, Appl1
C	45	72	6.6	536165	4	US-09-214-808-1	Sequence 1, Appl1

## ALIGNMENTS

RESULT 1  
US-08-816-241-2  
Sequence 2, Application US/08816241  
Patent No. 5804185  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Goll, Surya K.  
TITLE OF INVENTION: NOVEL RNA EDITING ENZYME  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/816,241  
FILING DATE: Filed Herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0239 US  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 2:











US-09-966-880A-8 (1-198) X US-09-040-482-2 (1-891)

QY 11 PheLeuTYGlnPheIysAsnValArgTrpAlaIysGlyArgGlyTrpTYrLeuCys 30  
 Db 306 TTTAAATCCAGTCCGGATGGATGGATGACCTTCCGGAGAACCAAGACCTTCCCTGCG 365  
 QY 31 TyrValValIysArgArgAspSerAlaThrSerPheSerLeuAspPheGlyTYrLeuArg 50  
 Db 366 TATGTGCTGAAGCACAAGGCAAGGGGGGCCAAGTCAGGATGTCGGGGATACCTAGAG 425  
 QY 51 AsnIysAsn---GlyCysHisValGlnLeuLeuPheLeuArgTYrIleSerAspTrpAsp 69  
 Db 426 GATAGACATCGCGCTGCCATGCAGAGGAAGCTTCTTCAACACCATCTGGCCA--GCC 482  
 QY 70 LeuAspProGlyArgCysTYrArgValIleTrpPheThrSerTrpSerProCysTYrAsp 89  
 Db 483 TTCACCCACCCCTGCGGTACATGTCACCTGTATGTCTCTCCAGCCCTGTGACGC 542  
 QY 90 CysAlaArgHisValAlaAspPheLeuArgGlyAsnProAsnLeuSerLeuArgIlePhe 109  
 Db 543 TGTCTCAACCGCATATACAAACCCCTTACCAAGACCAAGAACCTGCTGTCTCATCTCG 602  
 QY 110 ThrAlaArgLeuTYrPheCysGlnAspArgIysAlaGlnProGlu-----GlyLeu 126  
 Db 603 GTGGGTGACCTCTTCATCTGGAG-----GAGCCGGAATCCAGGCTGCTCG 650  
 QY 127 ArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThrPheIysAspTYrPheTYr 146  
 Db 651 AAGAAAGCTGAAGAGAGCGTGGCTGTAAACCTGGCATCATGAAAGCCCAAGACTTCAAAAT 710  
 QY 147 CysTrpPheMetTrpPheValGlnAsnHisGlu-----ArgThrPheGlyAlaTrpGlu 163  
 Db 711 GTCTGGCGAGATTTTGTGGAGCAAGAGAGGGGTGAATCCAAAGGCTTTCACACCTGGAG 770  
 QY 164 GlyLeuHisGlnIysAsnSerValArgLeuSerArgGlnLeuArgArgIle 179  
 Db 771 GACATTCAGGAACTTCTTATACTACGAGAGAAATTTGGCAGAGATC 818

```

? LENGTH: 879 base pairs
?
? TYPE: nucleic acid
?
? STANDEDNESS: single
?
? TOPOLOGY: linear
?
? MOLECULE TYPE: DNA (genomic)
?
? FEATURE:
?
? NAME/KEY: CDS
? LOCATION: 23..730
?
US-08-158-682A-3

```

US-09-966-880A-8 (1-198) X US-08-158-682A-3 (1-879

[illegible]



## ATTORNEY/AGENT INFORMATION:

NAME: Cooley, Ronald B.  
REGISTRATION NUMBER: 27,187  
REFERENCE/DOCKET NUMBER: ARCD:085  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 744-0090  
TELEFAX: (312) 245-4961  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 879 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)  
US-08-158-682A-1

## Alignment Scores:

Pred. No.:	2,89e-16	Length:	879
Score:	201.00	Matches:	56
Percent Similarity:	47.65%	Conservative:	25
Best Local Similarity:	32.94%	Mismatches:	61
Query Match:	18.51%	Indels:	28
		Gaps:	6

US-09-966-880A-8 (1-198) x US-08-158-682A-1 (1-879)

```

QY 24 ArgArgGluThrTyrLeuCyseTyrValValys-----Arg 35
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 137 CGGAAGAGACCTGCTGCTGATGAGATCAACTGGGAGAGGACACAGACTGCGGA 196
QY 36 ArgAspSerAlaThrSerPheSerLeuAspPheGlyTyrLeuArgAsnIysVal 55
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 197 CACACGAGCAAAACACC-----AACAA----- 220
QY 56 HisValGluLeuLeuPheLeu---ArgTyrIleSerAspTyrAspLeuAspProGlyArg 74
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 221 CACGTTGAAGTCATTTCATAGAAAATTACTACAGAAAGATCTTTGTCACAAACACC 280
QY 75 CysTyrArgValThrTyrPheThrSerTyrSerProCysTyrAspCysAlaArgHisVal 94
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 281 AGATGCTCATCTACTGCTGCTGCTGAGTCCCTGCTGGGAGTCTCCAGGCGCAT 340
QY 95 AlaAspPheLeuArgGlyAsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyr 114
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 341 ACAGATTGTTGAGCGCATACCCCATGTAACCTGTTATTATATAGACACGCGCTTAT 400
QY 115 PheCysGluAspArgGlyAlaGluProGluGlyLeuArgArgLeuHisAlaGlyVal 134
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 401 CACACGAGCATCTCGA---AATCGGCAAGACATCGAGCTTATTAGCAGCGGT 457
QY 135 GlnIleAlaIleMetThrPheIysAspTyrPheTyrCysTyrAsnThrPheValGluAsn 154
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 458 ACTATCCAGATCATGACGAGCAGACAGCTGCTGCTGAGGAATTTGTCAACTAC 517
QY 155 HisGluArgThrPheIysAlaIleTyrGluGlyLeuHisGluAsnSerValArgLeuSerArg 174
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 518 TCCCTTGAGATGAGCTCATTTGGCCAGATACCCCATCTGTGGGTGAGGCTG----- 571
QY 175 GlnLeuArgArgIleLeuLeuProLeuTyr 184
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 572 -----TACGTAAGTGAAGTCTAC 589

```

## RESULT 7

US-08-015-203-1  
Sequence 1, Application US/08015203  
Patent No. 5550034  
GENERAL INFORMATION:  
APPLICANT: Teng, Babie  
APPLICANT: Davidson, Nicholas O.  
APPLICANT: Barrant, Charles F.  
TITLE OF INVENTION: Apolipoprotein B RNA Editing Protein:  
TITLE OF INVENTION: Composition and Method  
NUMBER OF SEQUENCES: 2

## CORRESPONDENCE ADDRESS:

ADDRESSES: ARNOLD, WHITE & DURKEE  
STREET: 321 NO. 5550034th Clark Street, Suite 800  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60610  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentia Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/015,203  
FILING DATE: 19930209  
CLASSIFICATION: 435

## ATTORNEY/AGENT INFORMATION:

NAME: Cooley, Ronald B.  
REGISTRATION NUMBER: 27,187  
REFERENCE/DOCKET NUMBER: ARCD:069  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 744-0090  
TELEFAX: (312) 245-4961  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 879 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-015-203-1

## Alignment Scores:

Pred. No.:	2,89e-16	Length:	879
Score:	201.00	Matches:	56
Percent Similarity:	47.65%	Conservative:	25
Best Local Similarity:	32.94%	Mismatches:	61
Query Match:	18.51%	Indels:	28
		Gaps:	6

US-09-966-880A-8 (1-198) x US-08-015-203-1 (1-879)

```

QY 24 ArgArgGluThrTyrLeuCyseTyrValValys-----Arg 35
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 137 CGGAAGAGACCTGCTGCTGATGAGATCAACTGGGAGAGGACACAGACTGCGGA 196
QY 36 ArgAspSerAlaThrSerPheSerLeuAspPheGlyTyrLeuArgAsnIysVal 55
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 197 CACACGAGCAAAACACC-----AACAA----- 220
QY 56 HisValGluLeuLeuPheLeu---ArgTyrIleSerAspTyrAspLeuAspProGlyArg 74
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 221 CACGTTGAAGTCATTTCATAGAAAATTACTACAGAAAGATCTTTGTCACAAACACC 280
QY 75 CysTyrArgValThrTyrPheThrSerTyrSerProCysTyrAspCysAlaArgHisVal 94
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 281 AGATGCTCATCTACTGCTGCTGAGTCCCTGCTGGGAGTCTCCAGGCGCAT 340
QY 95 AlaAspPheLeuArgGlyAsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyr 114
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 341 ACAGATTGTTGAGCGCATACCCCATGTAACCTGTTATTATATAGACACGCGCTTAT 400
QY 115 PheCysGluAspArgGlyAlaGluProGluGlyLeuArgArgLeuHisAlaGlyVal 134
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 401 CACACGAGCATCTCGA---AATCGGCAAGACATCGAGCTTATTAGCAGCGGT 457
QY 135 GlnIleAlaIleMetThrPheIysAspTyrPheTyrCysTyrAsnThrPheValGluAsn 154
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 458 ACTATCCAGATCATGACGAGCAGACAGCTGCTGCTGAGGAATTTGTCAACTAC 517
QY 155 HisGluArgThrPheIysAlaIleTyrGluGlyLeuHisGluAsnSerValArgLeuSerArg 174
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 518 TCCCTTGAGATGAGCTCATTTGGCCAGATACCCCATCTGTGGGTGAGGCTG----- 571

```



QY 175 GlnleuArgArgIleuLeuProLeuTyr 184  
Db 572 -----TAGCTACTGGAACCTCTAC 589

RESULT 8

US-09-068-655-10/c  
Sequence 10, Application US/09068655A  
Patent No. 6136579  
GENERAL INFORMATION:  
APPLICANT: JACKSON, Ronald James  
TITLE OF INVENTION: METHOD OF PRODUCING ALPHA 2,3-SIALYLTRANSFERASE  
FILE REFERENCE: 026579-186  
CURRENT FILING DATE: 1998-08-31  
EARLIER FILING DATE: 1995-11-15  
EARLIER APPLICATION NUMBER: AU PN 6587  
EARLIER FILING DATE: 1996-11-15  
NUMBER OF SEQ ID NOS: 13  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 10  
LENGTH: 1806  
TYPE: DNA  
ORGANISM: Myxoma virus  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (2)..(439)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1405)..(1806)  
US-09-068-655-10

Alignment Scores:  
Pred. No.: 0.339 Length: 1806  
Score: 87.50 Matches: 38  
Percent Similarity: 41.07% Conservative: 8  
Best Local Similarity: 33.93% Mismatches: 43  
Query Match: 8.06% Indels: 23  
DB: 3 Gaps: 7

US-09-966-880A-8 (1-198) x US-09-068-655-10 (1-1806)  
QY 20 TPpAlaYsGIyArGArgGluThrTyLeuCyStyValVallySArGArgAspSerAla 35  
Db 888 TCGTCCGCTGCAAGCCGCCGACCTATACGTGCTATATACATGCTTACGAACAACGAC 829  
QY 40 ThSerPheSerLeuAspPheGlyTyLeuArgAnlySArGlyCySHIeValGluLeu 59  
Db 828 GCGTCGACCTGCTGATTTTGAAGAAACCGCTCGACGTCGGCGCGGCTCGGCGT 769  
QY 60 LeuPheLeuArgTyrlIleSerAsp-----TriAspLeuAsp 71  
Db 768 TCGTACGAATCCCTACATCCGAATATACGTACGATCGCGCTGCGTATACAGACACGT 709  
QY 72 ProGlyArgCyStyArGValThrTrpPheThrSerTrpSer-----ProCySty 88  
Db 708 CCGAAGCTTGACCGCTTCTACAGATG-----GAATGTCGCGCTCGTGAACCGGCTTAC 655  
QY 89 AspCyAlaArgHISValAlaAspPheLeuArg-----GlyAsnPro----- 102  
Db 654 AGCTGGCCAGG-----GCGTACGATTAACGGGATTCGGGTACCCCAATGGGGGCGCTA 601  
QY 103 ---AsnLeuSerLeuArgIlePheThrAlaArgLeu 113  
Db 600 ACCAATTATCCATT---ATTACACGGCTATATCTA 568

RESULT 9  
US-09-068-655-3/c  
Sequence 3, Application US/09068655A  
Patent No. 6136579  
GENERAL INFORMATION:  
APPLICANT: JACKSON, Ronald James

TITLE OF INVENTION: METHOD OF PRODUCING ALPHA 2,3-SIALYLTRANSFERASE  
FILE REFERENCE: 026579-186  
CURRENT APPLICATION NUMBER: US/09/068,655A  
CURRENT FILING DATE: 1998-08-31  
EARLIER FILING DATE: 1995-11-15  
EARLIER APPLICATION NUMBER: AU PN 6587  
EARLIER FILING DATE: 1996-11-15  
NUMBER OF SEQ ID NOS: 13  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 3  
LENGTH: 4713  
TYPE: DNA  
ORGANISM: Myxoma virus  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(549)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1515)..(2081)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (2084)..(3745)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (3776)..(4434)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (4440)..(4712)  
US-09-068-655-3

Alignment Scores:  
Pred. No.: 1.44 Length: 4713  
Score: 87.50 Matches: 38  
Percent Similarity: 41.07% Conservative: 8  
Best Local Similarity: 33.93% Mismatches: 43  
Query Match: 8.06% Indels: 23  
DB: 3 Gaps: 7

US-09-966-880A-8 (1-198) x US-09-068-655-3 (1-4713)  
QY 20 TPpAlaYsGIyArGArgGluThrTyLeuCyStyValVallySArGArgAspSerAla 35  
Db 998 TCGTCCGCTGCAAGCCGCCGACCTATACGTGCTATATACATGCTTACGAACAACGAC 939  
QY 40 ThSerPheSerLeuAspPheGlyTyLeuArgAnlySArGlyCySHIeValGluLeu 59  
Db 938 GCGTCGACCTGCTGATTTTGAAGAAACCGCTCGACGTCGGCGCGGCTCGGCGT 879  
QY 60 LeuPheLeuArgTyrlIleSerAsp-----TriAspLeuAsp 71  
Db 878 TCGTACGAATCCCTACATCCGAATATACGTACGATCGCGCTGCGTATACAGACACGT 819  
QY 72 ProGlyArgCyStyArGValThrTrpPheThrSerTrpSer-----ProCySty 88  
Db 818 CCGAAGCTTGACCGCTTCTACAGATG-----GAATGTCGCGCTCGTGAACCGGCTTAC 765  
QY 89 AspCyAlaArgHISValAlaAspPheLeuArg-----GlyAsnPro----- 102  
Db 764 AGCTGGCCAGG-----GCGTACGATTAACGGGATTCGGGTACCCCAATGGGGGCGCTA 711  
QY 103 ---AsnLeuSerLeuArgIlePheThrAlaArgLeu 113  
Db 710 ACCAATTATCCATT---ATTACACGGCTATATCTA 678

RESULT 10  
US-09-198-452A-1/c  
Sequence 1, Application US/09198452A  
Patent No. 6559294  
GENERAL INFORMATION:  
APPLICANT: Griffiths, R.  
TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragment thereof and uses thereof, in particular for the diagnosis, pre



	TITLE OF INVENTION: and treatment of infection
	FILE REFERENCE: 9710-003-999
	CURRENT APPLICATION NUMBER: US/09/198,452A
	CURRENT FILING DATE: 1998-11-24
	NUMBER OF SEQ ID NOS: 6849
	SEQ ID NO 1
	LENGTH: 1330025
	TYPE: DNA
	ORGANISM: Chlamydia pneumoniae
	FEATURE:
	NAME/KEY: misc_feature
	LOCATION: (1)..(15000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (15001)..(30000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (30001)..(45000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (45001)..(60000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (60001)..(75000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (75001)..(90000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (90001)..(105000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (105001)..(120000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (120001)..(135000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (135001)..(150000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (150001)..(165000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (165001)..(180000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (180001)..(195000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (195001)..(210000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (210001)..(225000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (225001)..(240000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (240001)..(255000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (255001)..(270000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (270001)..(285000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (285001)..(300000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (300001)..(315000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (315001)..(330000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (330001)..(345000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (345001)..(360000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (360001)..(375000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (375001)..(390000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (390001)..(405000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (405001)..(420000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (420001)..(435000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (435001)..(450000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (450001)..(465000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (465001)..(480000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (480001)..(495000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (495001)..(510000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (510001)..(525000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (525001)..(540000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (540001)..(555000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (555001)..(570000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (570001)..(585000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (585001)..(600000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (600001)..(615000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (615001)..(630000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (630001)..(645000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (645001)..(660000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (660001)..(675000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (675001)..(690000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (690001)..(705000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (705001)..(720000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (720001)..(735000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (735001)..(750000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (750001)..(765000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (765001)..(780000)
	OTHER INFORMATION: n=a or c or g or t
	NAME/KEY: misc_feature
	LOCATION: (780001)..(795000)
	OTHER INFORMATION: n=a or c or g or t
	NAME



```

LOCATION: (675001)..(690000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (690001)..(705000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (705001)..(720000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (720001)..(735000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (735001)..(750000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (750001)..(765000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (765001)..(780000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (780001)..(795000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (795001)..(810000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (810001)..(825000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (825001)..(840000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (840001)..(855000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (855001)..(870000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (870001)..(885000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (885001)..(900000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (900001)..(915000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature

```

## Alignment Scores:

```

Pred. No.: 9.63e+03 Length: 1230025
Score: 86.00 Matches: 55
Percent Similarity: 36.89% Conservative: 21
Best Local Similarity: 26.70% Mismatches: 57
Query Match: 7.92% Indels: 73
DB: 4 Gaps: 10

```

US-09-966-880A-8 (1-198) x US-09-198-452A-1 (1-1230025)

```

QY 9 ArglySphe---LeuTyGlnPheIysAsnValArgTPaIaIySgIyArgArgIuThr 27
DB 289205 AGAAATATCCTTATATCAAGAACCCACATGAGTCAATTAATCCGTAAGAAATCAG 289146
QY 28 TyIleuCyTyVal----- 32
DB 289145 ATCCCTTATACATCGTTGAGGGCGGTGAGCCGCCGCTGACCAATCTAACATCA 289086
QY 33 -----ValIyArgArgAspSerAlaThrSerPheSerLeuAspPheGlyTyIleuArg 50
DB 289085 AGCCTGTAAAGATTATTAACATCAATCCATTAATTTCTTGTGTGCTTGCTTAATTA--- 289029
QY 51 AsnIyAsnGlyCyshiValIuIleuPheIeuArgTyIle-SerAspTPaAspLe 70

```

```

DB 289028 -----TTACGATNCTTACGAGCTTATATG 289002
QY 70 uAspProGlyArgCyTyArgValThrTPheThrSerTPSerProCyTyAsp-- 89
DB 289001 TAGGCTCT-----TGTTTTGCTCTATTTTATATCTGTGAGCTCTGCTGTATGA 288948
QY 90 -----CysAlaArgHisValAlaAspPheIeuArgIyAsnPr 102
DB 288947 CCGGACGATATATGATGCTGTGCTGACGAGG-----AACAGGATTC 288903
QY 102 AsnIeuSerIeuArgIlePheThrAlaArgLeuTyPheCySgIuAspArgIyAlaG1 122
DB 288902 TCCTCTGTAAATAGGCTTTTACTTGC-----TTCTGTATATAGCGTGCACGAAA 288852
QY 122 uPro-GluGlyIleuArgArgIleuHisArgAlaGlyValGlnIleAlaIle-MetTPRpe 141
DB 288851 GCCAAGAGGTGTTTCATCATATGATCCGCACTATTCGTTGTATAGTGAACA--- 288795
QY 142 LysAspTyPheTyCySTpAntThrPheValGluAsnHisGluArgThrPheIyAla 161
DB 288794 ---CAATATTCAGATGC---CTTACTTCTTATCACTCCAGAA----- 288756
QY 162 TrpGluGlyIleuHisGluAsnSerValArgIeuSerArgIleuArgArgIleuIeu 181
DB 288755 -----CTAAAAAAGTGTACTTC 288738
QY 182 ProIeuTyArgIu 185
DB 288737 AGTCTATCGAA 288726

RESULT 11
US-09-107-532A-2790
Sequence 2790, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: LYNN A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESS: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Arinello, Pamela Deneka
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 2790:
SEQUENCE CHARACTERISTICS:
LENGTH: 1356 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)

```



```

?      HYPOTHETICAL: NO
?      ANTI-SENSE: NO
?      ORIGINAL SOURCE:
?      ORGANISM: Enterococcus faecium
?      FEATURE:
?      NAME/KEY: misc feature
?      LOCATION: (B) LOCATION 1..1356
?      SEQUENCE DESCRIPTION: SEQ ID NO: 2790:
?      US-09-107-532A-2790

```

Alignment Scores:	
Pred. No.:	1.75
Score:	80.50
Percent Similarity:	37.97%
Best local Similarity:	24.68%
Query Match:	7.41%
DB:	4
Length:	1356
Matches:	39
Conservative:	21
Mismatches:	59
Indels:	39
Gaps:	8

US-09-966-880A-8 (1-198) X US-09-107-532A-2790 (1-1356)

QY 45 LeuYhgrslnlykshsnqlyCyshisvalgluileuleupheliuAsGTYrllleSerAspTrp 68  
DB 433 CTGATTGACGAAGACGGCGCAACGCTTAAAGAAATCATTTTACGCTACGCAAGAAAAATGG 492  
QY 69 AspLeuAspProGlyArgCysTyArgValIhTrpPheThrSerTrpSerProCysTy 88  
DB 493 GATTAGACCTGCT-----TTGTTGATGGTGTAGAAAAAGAA 531  
QY 89 Asp-----CysAlaArgHisValAlaAspPheLeuArgGlyAsnProAsnLeuSerLeu 106  
DB 532 AATGTTCCTGGTGTAGTTAGTTGATGCGATCGCAAGTATCTCGCGAAGATGCA 591  
QY 107 ArgIlePheThrAlaArgLeuTyPheCysGluAspArg-----LysAlaGluPro 123  
DB 592 GGTCTTTTCATCAACGCGCATACTCT---GAAGATCAATGATGCTGAAGCAAGAACCA 648  
QY 124 -----GluGlyLeuArgArgLeuHis-----Arg 131  
DB 649 TTCATGCTTTGGGTGATCGAAGCGGCTTAAGAATGGAGATCAATCACTCAACAG 708  
QY 132 AlaGlyValGlnIleAlaIleMetCThrPheLysAspTyPheTyCysTrpAsnThrPhe 151  
DB 709 GCTGGCTTGAATGCATCATCAACA-----GATGATATG 741  
QY 152 ValGluAsnHisGluArgThrPheLysValaTrpGluGlyLeuHisGluAsnSerValaArg 171  
DB 742 ACTGCTTAATGCTGAACGTAAAGTGATGCTTAATGAACCGCATACAGCATGGTCCCT 801  
QY 172 LeuSerArgGlnLeuArgArgIleLeuLeuProLeuTyGluValAspAspLeu 189  
DB 802 CTTCCT-----TTGTGGCGGAATTTGAAACTGTATGAAAAAGTA 840

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```

1      APPLICATION NUMBER: US/09/221,017B
2      FILING DATE: 23-DEC-1998
3      CLASSIFICATION:
4      PRIOR APPLICATION DATA:
5      APPLICATION NUMBER: PP1182
6      FILING DATE: 31-DEC-1997
7      PRIOR APPLICATION DATA:
8      APPLICATION NUMBER: PP1546
9      FILING DATE: 30-JAN-1998
10     PRIOR APPLICATION DATA:
11     APPLICATION NUMBER: PP2911
12     FILING DATE: 09-APR-1998
13     PRIOR APPLICATION DATA:
14     APPLICATION NUMBER: PCT/AU98/01023
15     FILING DATE: 10-DEC-1998
16     ATTORNEY/AGENT INFORMATION:
17     NAME: MONROY, Gladys H
18     REGISTRATION NUMBER: 32,430
19     REFERENCE/DOCKET NUMBER: 27340-020021,000
20     TELECOMMUNICATION INFORMATION:
21     TELEPHONE: 650-813-5600
22     TELEFAX: 650-494-0792
23     TELEX: 706141
24     INFORMATION FOR SEQ ID NO: 123:
25     SEQUENCE CHARACTERISTICS:
26     LENGTH: 7077 base pairs
27     TYPE: nucleic acid
28     STRANDEDNESS: double
29     TOPOLOGY: circular
30     MOLECULE TYPE: DNA (genomic)
31     HYPOTHEetical: NO
32     ANTI-SENSE: UNKNOWN
33     ORIGINAL SOURCE:
34     ORGANISM: PORPHYROMONAS GINGIVALIS
35     FEATURE:
36     NAME/KEY: misc_feature
37     LOCATION: 1...7077
38     US-09-221-017B-123

```

Alignment Scores:		
Pred. No.:	24.4	7077
Score:	80.00	22
Percent Similarity:	47.628	
Best Local Similarity:	34.928	17
Query Match:	7.378	16
DB:	4	3
US-03-966-880A-8 (1-198)	x	US-09-221-017B-123 (1-7077)

Oy		4	LeuLeuMetAsnArgGlyGlyPheLeuTyrClnPhelyshsnValAlaGTPAlaLysGly	23
Dd		3900	TTGTATATTAAATCCAAAACGTTTTATATATAGATCGTCACCGCTTGATGGGAAGAAC	3959
Oy		24	ArgArgIunThrTyrLeuCysTyrValValLys-----ArgArgAspSerAlaThr	40
Dd		3960	GCCGAGGAATCTTTCTTGAAAGCATAAAAAGCAAGCACCCGCCGAAGTCTCTG---	4016
Oy		41	SerPheSerLeuAspPheGlyTyrLeuArgAsnLysasnGlyCysHisValGluleuLeu	60
Dd		4017	-----CGAAGCTCCAGCGGGTGC-----CTTGCT	4040
Oy		61	PheLeuArg	63
Dd		4041	TTTTGTGCT	4049

RESULT 13  
 US-09-252-991A-7743/C  
 ; Sequence 7743, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS







Db	413	CGATGCGAGCGCTTACCTGATCTGCAACGCGCTACGACGCGCAATCCCGGC	354
Qy	178	GILEuleuProleuTyrgluValAspAspleu---ArgAspAlaPheargThrleuGI	197
Db	353	AGGCCAACTACCGGCGGATCCAGTACGACACCTTATGCGTATGCTGCCCAAGCCGAGG	294
Qy	197	Y 197	
Db	293	C 293	

Search completed: March 13, 2004, 14:53:23  
 Job time : 328 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: March 13, 2004, 10:11:37 ; Search time 275 Seconds  
(without alignments)  
2650.750 Million cell updates/sec

Title: US-09-966-880A-8  
Perfect score: 1086  
Sequence: 1 MDSLLMNRKFLYQFRTYRW.....ILLPLVVDLRLDAFRLGL 198

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2432557 seqs, 1847978884 residues  
Total number of hits satisfying chosen parameters: 4865114

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n model -DEV=xlh  
-DB=publshdApplications\_NA -QMT=fastap -SUFFIX=mpb -MINMATCH=0.1  
-LOOPEXT=0 -LOOPEXT=0 -UNITS=Dits -START=1 -END=1 -MATRIX=Ditsum62  
-TRAN=human40.cdi -LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09966880 @CGN 1.1.164 @runat 11032004 172743\_26651  
-NCPU=6 -ICRT=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100  
-LONGLOG -DEV=TIMEOUT=120 -WARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELXT=7

Database : Published Applications NA:

1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PTCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PTCT\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/2/pubpna/US10D\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/2/pubpna/US10E\_NEW\_PUB.seq.\*  
17: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
18: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result	Score	Query Match	Length	ID	Description
--------	-------	-------------	--------	----	-------------

1	1086	100.0	2818	9	US-09-966-880A-7	Sequence 7, Appl
2	1008	92.8	2440	9	US-09-966-880A-1	Sequence 10, Appl
3	644.5	59.3	6664	9	US-09-966-880A-10	Sequence 1, Appl
4	644.5	59.3	11204	9	US-09-966-880A-35	Sequence 35, Appl
5	505	46.5	271	9	US-09-966-880A-13	Sequence 13, Appl
6	390	35.9	1155	15	US-10-460-923-1	Sequence 1, Appl
7	369.5	34.0	1534	9	US-09-925-300-699	Sequence 173, App
8	369.5	34.0	987	9	US-09-925-300-699	Sequence 699, App
9	363.5	34.0	1143	13	US-10-098-841-268	Sequence 268, App
10	363.5	33.5	1348	9	US-09-880-107-3282	Sequence 3282, App
11	348	32.0	883	14	US-10-247-671-62	Sequence 62, Appl
12	343	31.6	1008	15	US-10-131-827-8890	Sequence 8890, App
13	308.5	28.4	2151	9	US-09-822-830A-359	Sequence 359, App
14	262	24.1	201143	12	US-10-240-425-1099	Sequence 1099, App
15	260	23.9	148	9	US-09-966-880A-12	Sequence 12, Appl
16	257	23.7	371	10	US-09-918-995-30237	Sequence 30237, A
17	249.5	23.0	2913	15	US-10-027-632-112023	Sequence 112023,
18	249.5	23.0	2913	15	US-10-027-632-112024	Sequence 112024
19	238.5	22.0	1404	14	US-10-198-846-13472	Sequence 13472, A
20	218.5	20.1	879	14	US-10-157-031-13	Sequence 13, Appl
21	212	19.5	116	9	US-09-866-880A-14	Sequence 14, Appl
22	209	19.2	2773	15	US-10-104-047-1759	Sequence 1759, App
23	206	19.0	390	9	US-09-796-692-9051	Sequence 9051, App
24	206	19.0	390	14	US-10-040-862-9051	Sequence 9051, App
25	206	19.0	390	15	US-10-057-475B-9051	Sequence 9051, App
26	206	19.0	390	15	US-10-154-884B-9051	Sequence 9051, App
27	198	18.2	821	15	US-10-388-934-189	Sequence 189, App
28	192	17.7	674	14	US-10-257-826A-150	Sequence 150, App
29	175	16.1	556	9	US-09-796-692-45061	Sequence 5061, App
30	175	16.1	556	14	US-10-040-862-5061	Sequence 5061, App
31	175	16.1	556	15	US-10-057-475B-5061	Sequence 5061, App
32	175	16.1	556	15	US-10-154-884B-5061	Sequence 5061, App
33	167	15.4	445	15	US-10-378-029-25	Sequence 25, Appl
34	166.5	15.3	476	10	US-09-918-995-33385	Sequence 3385, A
35	164	15.1	553	9	US-09-796-692-4504	Sequence 4504, App
36	164	15.1	553	14	US-10-040-862-4504	Sequence 4504, App
37	164	15.1	553	15	US-10-057-475B-4504	Sequence 4504, App
38	164	15.1	553	15	US-10-154-884B-4504	Sequence 4504, App
39	162.5	15.0	572	9	US-09-864-761-9416	Sequence 9416, App
40	157.5	14.5	328	9	US-09-796-692-5737	Sequence 5737, App
41	157.5	14.5	328	14	US-10-040-862-5737	Sequence 5737, App
42	157.5	14.5	328	15	US-10-057-475B-5737	Sequence 5737, App
43	157.5	14.5	328	15	US-10-154-884B-5737	Sequence 5737, App
44	154	14.2	456	14	US-10-029-386-25187	Sequence 25187, A
45	122	11.2	505	9	US-09-796-692-4443	Sequence 4443, App

## ALIGNMENTS

RESULT 1  
US-09-966-880A-7  
Sequence 7, Application US/09966880A  
Patent No. US20020164743A1  
GENERAL INFORMATION:  
APPLICANT: Homjo, Tasuku  
APPLICANT: Muramatsu, Masamichi  
TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
FILE REFERENCE: 06501-088001  
CURRENT APPLICATION NUMBER: US/09/966, 880A  
CURRENT FILING DATE: 2001-09-28  
PRIOR APPLICATION NUMBER: PCT/JP00/01918  
PRIOR FILING DATE: 2000-03-28  
PRIOR APPLICATION NUMBER: JP 11-371382  
PRIOR FILING DATE: 1999-12-27  
PRIOR APPLICATION NUMBER: JP 11-178999  
PRIOR FILING DATE: 1999-06-24  
PRIOR APPLICATION NUMBER: JP 11-87192  
PRIOR FILING DATE: 1999-03-29  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: PatsSeq for Windows Version 4.0  
SEQ ID NO 7  
LENGTH: 2818  
TYPE: DNA



ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (80)...(673)  
 FEATURE:  
 NAME/KEY: 5'UTR  
 LOCATION: (1)...(79)  
 FEATURE:  
 NAME/KEY: 3'UTR  
 LOCATION: (677)...(2818)  
 US-09-966-880A-7

Alignment Scores:  
 Pred. No.: 2,93e-140 Length: 2818  
 Score: 1086.00 Matches: 198  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 Gaps: 0

US-09-966-880A-8 (1-198) x US-09-966-880A-7 (1-2818)

QY 1 MetAapSerLeuMetAapSerLeuTyrglnPheLysAsnValArgTTP 20  
 DB ATGACAGCCTCTGATGACCGGAGAGTTCTTACCAATTCAAAATGTCGCTGG 139  
 QY 21 AAlaySgLYARgAGluThrTyrlencysTyrlvalVallvSARgASpSerAlaThr 40  
 DB GCTAAGGCTCGCGTGAAGCCTGCTGCTACGTAAGAGAGCGCTGACGCTACA 199  
 QY 41 SerPheSerLeuAapPheGlyTyrlencysAsnLysAsnGlyCySHSvalGluLeu 60  
 DB TCCTTTTCACTGACCTTGGTTGTTATCTTGCATAAAGAGCGCTGCCAGTGGATTGCTC 259  
 QY 61 PheLeuArgTyrlleSerAapTTPAapLeuAapProGlyArgCySlyrArgValThrTTP 80  
 DB TTCCTCCGCTACATCTCGAGACTGCGAGCCTAGACCTGCGCCCTGCTACCGCGTCACTCG 319  
 QY 81 PheThrSerTTPSerProCySlyrAapCySalyaArgHisValAlaAapPheLeuArgGly 100  
 DB TTCCACCTCTCGAGCGCCCTGCTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 379  
 QY 101 AsnProAenLeuSerLeuArgTlePheThrAlaArgLeuTyrlPheCySgluAapArgLys 120  
 DB AACCCCAACCTCAAGTGTGAGATCTTCAACCGCGCCCTCTACTTCTGTGAAGACCGCAG 439  
 QY 121 AAlayProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140  
 DB GGTGAGCCGAGAGGCGCTGCGCGCTGACCGCGCGGAGTGAATAGCATCATGAC 499  
 QY 141 PheLysAapTyrlPheTyrcySTTPAenThrPheValGluAsnHisGluArgThrPheLys 160  
 DB TTCAAAGATTATTTTACTGCTGCAATCTTTTGTGAAGAACCAAGAAAGAACTTTCAA 559  
 QY 161 AAlaTTPGluGlyLeuHisGluAsnSerValArgLysSerArgGluLeuArgArgLys 180  
 DB GCCTGGAGAGGCTGATAAATTCAGTTCCTCTCCAGACACTTCGCGCATCTT 619  
 QY 181 LeuProLeuTyrglnValAspAapLeuArgAapAlaPheArgThrLeuGlyLeu 198  
 DB TTGCCCTGTATGAGGTTGATGACTTACAGAGAGCATTTGTACTTTGGGACTT 673

RESULT 2  
 US-09-966-880A-1

Sequence 1, Application US/09966880A  
 Patent No. US20020164743A1

GENERAL INFORMATION:  
 APPLICANT: Honjo, Tasuku  
 APPLICANT: Muramatsu, Masamichi  
 TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
 FILE REFERENCE: 06501-088001  
 CURRENT APPLICATION NUMBER: US/09/966, 880A  
 CURRENT FILING DATE: 2001-09-28

PRIOR APPLICATION NUMBER: PCT/JP00/01918  
 PRIOR FILING DATE: 2000-03-28  
 PRIOR APPLICATION NUMBER: JP 11-371382  
 PRIOR FILING DATE: 1999-12-27  
 PRIOR APPLICATION NUMBER: JP 11-178999  
 PRIOR FILING DATE: 1999-06-24  
 PRIOR APPLICATION NUMBER: JP 11-87192  
 PRIOR FILING DATE: 1999-03-29  
 NUMBER OF SEQ ID NOS: 36  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 1  
 LENGTH: 2440  
 TYPE: DNA  
 ORGANISM: Mus musculus  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: (93)...(686)  
 FEATURE:  
 NAME/KEY: 5'UTR  
 LOCATION: (1)...(92)  
 FEATURE:  
 NAME/KEY: 3'UTR  
 LOCATION: (690)...(2440)  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: (1)...(2440)  
 OTHER INFORMATION: n = A,T,C or G  
 US-09-966-880A-1

Alignment Scores:  
 Pred. No.: 1.76e-129 Length: 2440  
 Score: 1008.00 Matches: 183  
 Percent Similarity: 95.94% Conservative: 6  
 Best Local Similarity: 92.89% Mismatches: 8  
 Query Match: 92.82% Indels: 0  
 Gaps: 0

US-09-966-880A-8 (1-198) x US-09-966-880A-1 (1-2440)

QY 1 MetAapSerLeuMetAapSerLeuTyrglnPheLysAsnValArgTTP 20  
 DB ATGACAGCCTCTGATGACCGGAGAGTTCTTACCAATTCAAAATGTCGCTGG 152  
 QY 21 AAlaySgLYARgAGluThrTyrlencysTyrlvalVallvSARgASpSerAlaThr 40  
 DB GCCAAGAGAGCGGATGACGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 212  
 QY 41 SerPheSerLeuAapPheGlyTyrlencysAsnLysAsnGlyCySHSvalGluLeu 60  
 DB TTCCTGCTACTGACCTTGGCCACCTTGCACCAAGTCTGCTGCTGCTGCTGCTGCTGCTG 272  
 QY 61 PheLeuArgTyrlleSerAapTTPAapLeuAapProGlyArgCySlyrArgValThrTTP 80  
 DB TTCTTAGCTACATCTCAAGCTGAGACTGAGCCGCGCGCTGTTACCGGCTCACTGG 332  
 QY 81 PheThrSerTTPSerProCySlyrAapCySalyaArgHisValAlaAapPheLeuArgGly 100  
 DB TTCACTCTCGAGCGCCGCTGCTACGCTGACCGCGCGCTGCTGCTGCTGCTGCTGCTGCTG 392  
 QY 101 AsnProAenLeuSerLeuArgTlePheThrAlaArgLeuTyrlPheCySgluAapArgLys 120  
 DB AACCCCAACCTCAAGTGTGAGATCTTCAACCGCGCCCTCTACTTCTGTGAAGACCGCAG 452  
 QY 121 AAlayProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140  
 DB GCTGAGCCTGAGAGGCTGCGAGACTGACCGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTG 512  
 QY 141 PheLysAapTyrlPheTyrcySTTPAenThrPheValGluAsnHisGluArgThrPheLys 160  
 DB TTCAAAGATTATTTTACTGCTGCAATCTTTTGTGAAGAACCAAGAAAGAACTTTCAA 572  
 QY 161 AAlaTTPGluGlyLeuHisGluAsnSerValArgLysSerArgGluLeuArgArgLys 180







Pred. No.:	8,85e-78	Length:	11204
Score:	644.50	Matches:	177
Percent Similarity:	24.02%	Conservative:	0
Best Local Similarity:	24.02%	Mismatches:	2
Query Match:	59.35%	Indels:	560
DB:	9	Gaps:	2

US-09-966-880A-8 (1-198) x US-09-966-880A-35 (1-11204)

QY	3	SeIeuLeuWeLksnrgahryglsPhleuYrGlnPheLysAsnValArTrpaLaLys	22
Dd	6278	AGCCTCTTGATGAACCGAGAAGATTCTTTACCAATTAAAAGTCCGGCTGGCTAAG	6337
OY	23	GLArghrglJunhrlYrLeuCysTyriValIvalySarGrASpSerAlaThrSerPhe	42
Dd	6338	GGTCGGGTAGACCTAACCTGCTGCTACGTAGTGAAGACGGCGACAGTGTCTACATCTTT	6397
OY	43	SerLeuApPheGlYTyrieuArGaLnLyS-----	52
Dd	6398	TCACTGGACTTTGGTTATCTTCGCCAAATAA--GGATCATTAATAAGTCAGCTTGCAGACAG	6456
OY	52	-----	52
Dd	6457	TTTTAATGTCACACTGTGAGTGTCTTTTAGAGCCACCCTGCTGATGATTACTTCATCTTT	6516
OY	52	-----	52
Dd	6517	TTTTGGCAATTGTGTCTCTATCACATTCCCAATCTTTTTTTTATTCTTTTCCATG	6576
OY	52	-----	52
Dd	6577	TCCATGACACCATATAGACATGGGCCAAAAATATGTATTATTCCTCCCAGTAATGC	6636
OY	52	-----	52
Dd	6637	TGGGACCCCTAATACACCTCTTCCTTCAGTGCAGAACACAACCTCTCCCAACTGTTA	6696
OY	52	-----	52
Dd	6637	CCAGCTTCCTACAGACTGTGAATGTCTTGAGTTATTTAAGCTAAAGCATTTTATA	6756
OY	52	-----	52
Dd	6757	TGGAGAAATATATCAGCTGTGCCAGCAAAATTTTAAATGTGAAAAACAATGTGTCTC	6816
OY	52	-----	52
Dd	6817	TTAAGCATTTTGAATAATTAAGAGAAGAAATTTGGAAAAAATTAACGGTGTCCAATT	6876
OY	52	-----	52
Dd	6877	CTGTTTCCAAATGATTTCTTTTCCCTCCTACTCACTGGGTGTAGGCGAGTAATACA	6936
OY	52	-----	52
Dd	6937	TTCAACATGTGTATCCCCAGAAAACTACAGAGACCTCGCTGATGATTAATTAATGA	6996
OY	52	-----	52
Dd	6997	TCCTTCGGCTACCCGAGAAATTACATTCCAGAGACTCTTTCACCAAAATCCAGATGG	7056
OY	52	-----	52
Dd	7057	GTTTACATAACTTCGCCCATGGGTATCTCTCTCTCTAACAAGCTGTGACGCTGGG	7116
OY	52	-----	52
Dd	7117	CTTGAGGAATCTCAGGGAAGCATCCGTGGGTGMAAGTCATGTCTGGCTGTGTTT	7176
OY	52	-----	52
Dd	7177	GAATGTTAATTAACATGCAATTTCTTTCCTCACTTTGATTAATGAATACATCCCAATCT	7236



QY	52	-----	52	-----	52
Db	7237	CCTTCCTATTGGTGACATGACACATCTCTATTTCAGAAAGCGTTGATTATTCAAGCACT	7236		
QY	52	-----	52	-----	52
Db	7297	TTCAATTACTCTCATGAGAGAGCTATTACTCTCTTACAAATACCATTCTGTCTCTT	7356		
QY	52	-----	52	-----	52
Db	7357	ACCAAAATCTATTTCCTCTTTAGATCTCCCAATGTCCTCATAAACTGTCTGCT	7416		
QY	52	-----	52	-----	52
Db	7417	CCACCTAGTGTCCAGGATATTTTCACAATGTTACATCAAGGCACTTACGCATT	7476		
QY	52	-----	52	-----	52
Db	7477	TCCTTGTCAAAAGGTGCAAAAAGCACTTCATTAAACAATAATTCTTGGTGAAGTA	7536		
QY	52	-----	52	-----	52
Db	7537	GTGTGATGCTGCTTCCTCCCACTGAGGCACTTGCTTCTCATTCCACAAAAACCA	7596		
QY	52	-----	52	-----	52
Db	7597	TAGCCTTCCTCACTGACAGGACTAGTGTCCAAAGGTTCAAGCTTACTAGTGT	7656		
QY	52	-----	52	-----	52
Db	7657	GCTCTTTGAGCAAGTCTTAGCTCTCTGTACACAGGACAAATAGCTGACAGATCC	7716		
QY	52	-----	52	-----	52
Db	7717	CCAAAGATCATTTGACAGAGACATGACTTAAGGCTACAGAGCGGCATTAAGTCAGTGA	7776		
QY	53	-----	53	-----	53
Db	7777	ATTTTAGCGTGTCTCTCTGTCTCTCCAGAACGGCTCCAGGTGAATGTCTCTCTC	7836		
QY	63	ArgTyrIleSerAspTyrAspLeuAspProGlyArgCysTyrArgValThrTyrPheThr	82		
Db	7837	CGCTACACTCGGACTGGGACCTTAACCTTGCGCTCTACCGGTCACTGCTTCAACC	7896		
QY	83	SeITyrSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgIleAsnPro	102		
Db	7897	TCCTGAGACCCCTGCTACACTGTGCCGACATGTGGCCGACTTCTCGAGGGAACCC	7956		
QY	103	AsnLeuSerLeuArgIlePheThrIleAlaArgLeuTyrPheCysGluAspArgIleAlaIle	122		
Db	7957	AACCTCACTCGAGGATCTTCAACCGCGGCTCTTACTGTGAAGACCGCAAGGCTTAG	8016		
QY	123	ProGluGlyLeuArgArgLeuHisArgIleArgIleValGlnIleAlaIleMetThrPheLys	142		
Db	8017	CCCGAGGGGCTGCGCGGCTGCACCGCGCGGGGTGCAAATAGCATATGACTTTCAA-	8075		
QY	142	-----	142	-----	142
Db	8076	AGGTGCAAAAGGCTTCCGGGACAGGCGAGTGCAGACCCGCACTTGGGATTTCGATG	8135		
QY	142	-----	142	-----	142
Db	8136	CGGAATGAATGATTAGTGGGGAAGCTGAGGGGAAAGATGGGGGGAATTCGTGTCA	8195		
QY	142	-----	142	-----	142
Db	8196	CCTTGGAGCCGAAATTAAAGATTGAAAGCAGAGAAAGAGTGAATGGCTCAGAGACAG	8255		
QY	142	-----	142	-----	142
Db	8256	GCCCCGAGGAATGAGAAAATGAGGCGACGGTGTCTTCCCTCGATTGGAACCTG	8315		
QY	143	-----	143	-----	143

```

Db      8316 AACGTCCTCTAATCCCCCAATCCCAGCCTTTTTCCTTTTTTCTTTTTTTTTTGAAAGATTAT 8377
Qy      145 PheTYrCySTIPAsnThr-PheValGIuaSnHisGIUarThrPheIysAlatTPGluNGly 164
Db      8376 TTtTAACGTGGGAAtACTTTTGTGAGAACAACAAGAGAACttTTCaaAGcCTCGGAgAGg 8435
Qy      165 LeuHISgluaSernValArgLeuSerThrGlInleuArgTrglIeuleu 181
Db      8436 CTGCATGAATAATTCAGTTGTCTCTCCAGACAGCgtTCGGCGCATCCtTTTG 8486

RESULT 5
US-09-966-880A-13
; Sequence 13, Application US/09966880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; APPLICANT: Muramatsu, Masamichi
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
FILE REFERENCE: 06501-088001
CURRENT FILING DATE: 2001-09-28
PRIOR APPLICATION NUMBER: PCT/JPO00/01918
PRIOR FILING DATE: 2000-03-28
PRIOR APPLICATION NUMBER: JP 11-371382
PRIOR FILING DATE: 1999-12-27
PRIOR APPLICATION NUMBER: JP 11-178999
PRIOR FILING DATE: 1999-06-24
PRIOR APPLICATION NUMBER: JP 11-87192
PRIOR FILING DATE: 1999-03-29
NUMBER OF SEQ ID NOS: 36
SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 13
LENGTH: 271
TYPE: DNA
ORGANISM: Homo sapiens
US-09-966-880A-13

Alignment Scores:
Pred. No.: 8,366-61 Length: 271
Score: 505.00 Matches: 90
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 46.50% Indels: 0
DB: Gaps: 0

US-09-966-880A-8 (1-198) x US-09-966-880A-13 (1-271)
Qy      53 AasGLICyNHISvalGIuLleuPheIueArgrYrtIIeserAprrPaapIeuarPro 72
Db      1 AAacggctgcAcagtgaaattgccttcctccgctaCatcggaactggaacctgaacct 60
Qy      73 GLIARCYETfARgVALtThrTPRphenSeITRPserProcSYIRASPcySaLaarg 92
Db      61 GGCGCGTGCTtaCGCGGTcaCTCGGTcaCTCCCGAGGCCCTGaCaGActGTGCCGA 120
Qy      93 HISvalIASpPhelEuARGLYASnpRoIsnuSERIEuARGIllePeTHraLaarg 112
Db      121 CATGTGGCCGaatTTCCTCTCGAGGAAACCACCACTCAGTCTGAGATCTTCAACCGCGCC 180
Qy      113 LeuTYrPhECySGLIuSPaRGLyVaLGluPrOclndIyeuARGaRguHuHSARGala 132
Db      181 CTTTaCTTGTGTGAGAGACCGCACAGGCTGaGCCAGGGGCGCTGAGCACCGCGCC 240
Qy      133 GIvAIglnlIALalIEMetThrPhElys 142
Db      241 GGGGTCAAATAGCCAATCATGACTTCAAA 270

RESULT 6
US-10-460-923-1
; Sequence 1, Application US/10460923
; Publication No. US20040009951A1
; GENERAL INFORMATION:
```



APPLICANT: MALIM, Michael H.  
APPLICANT: SHEEHY, Ann M.  
APPLICANT: HARRIS, Reuben S.  
APPLICANT: BISHOP, Kate N.  
APPLICANT: NEUBERGER, Michael S.  
APPLICANT: GADDIS, Nathan C.  
APPLICANT: SIMON, James H.M.  
TITLE OF INVENTION: DNA Deamination Mediates Innate Immunity to Retroviral Infection  
FILE REFERENCE: 22253-74380  
CURRENT APPLICATION NUMBER: US/10/460,923  
CURRENT FILING DATE: 2003-06-13  
PRIOR APPLICATION NUMBER: US 60/388,513  
PRIOR FILING DATE: 2002-06-13  
PRIOR APPLICATION NUMBER: US 60/472,952  
PRIOR FILING DATE: 2003-05-23  
NUMBER OF SEQ ID NOS: 12  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 1  
LENGTH: 1155  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-460-923-1  
  
Alignment Scores:  
Pred. No.: 8,36e-44 Length: 1155  
Score: 390.00 Matches: 83  
Percent Similarity: 61.62% Conservative: 31  
Best Local Similarity: 44.86% Mismatches: 59  
Query Match: 35.91% Indels: 12  
DB: Gaps: 3  
US-09-966-880A-8 (1-198) x US-10-460-923-1 (1-1155)  
  
QY 6 MetAsnAArgAlyGlyPheLeuTyrglnPheLyAsnValAArgTPAlaLySGlyAArgA 25  
DB 589 ATGATTCACCCACATTCATCTCACTTACAACTTACAACTTGGGTCAGAGAGCGCAT 648  
QY 26 GluThrTyLeuCyTyValValValysAArgAseAalathSerPheSerLeuasp 45  
DB 649 GAGACTTACCTGTGTATGAGGTGAGGAGCATGACATGACCTGGTCTGCTGAAC 708  
QY 46 -----PheGlyTyrLeuAArgAAsnLyS-----AsnGlyCyS 55  
DB 709 CAGCGCAGGGGCTTCTATGACACCGCTCCACATAAACAGGTTCTTGAAGCGGC 768  
QY 56 HisValGlnLeuLeuPheLeuAArgTyrlleSerAspTPAAspProGlyAArgCyS 75  
DB 769 CATGACAGGCTGTGCTCTGAGCGTGAATCCCTTTGGAAGCTGAGACCTGAGACCGAGAC 828  
QY 76 TyArgValThrTrpPheThrSerTrpSerProCyTyAAspCyValAArgHisValAla 95  
DB 829 TACAGGTTACCTGCTTCACTCCCTGAGCGCCCTGCTTCACTGAGCCCGCAATGCT 888  
QY 96 AspPheLeuAArgGlyAsnProAsnLeuSerLeuAArgIlePheThrAlaArgLeuTyrPhe 115  
DB 889 AATTCATTCAAAAACAAACACGTCGAGCTGTGACATCTTCACTGCGCCGCACTAT--- 945  
QY 116 CysGluAspAArgLyValAgluProGlnGlyLeuAArgLyHisAArgAlaGlyValGln 135  
DB 946 ---GATGATCAAGAGAGATGTCAGAGGGGCTGCCACCTGGCCGAGGCTGGAGCCCAA 1002  
QY 136 IleAlaIleMetThrPheLyAspTyrglnPheTyrcySTPAsnThrPheValGluAsnHis 155  
DB 1003 ATTCAAAATATGACATACGTAATTTAAGCACTCTGAGACCTTTGTGTGACACCAAG 1062  
QY 156 GluAArgThrPheLyAlaArgIleGlyLeuHisGluAsnSerValAArgLeuSerAArgGln 175  
DB 1063 GAGTGTCCCTTCAGAGCCCGGATGACATGATGACACAGCCAGACCTGATGGAGAG 1122  
QY 176 LeuAArgGlyIleLeu 180  
DB 1123 CTGCGGGCCATTTCTC 1137

RESULT 7  
US-09-729-674-173  
Sequence 173, Application US/09729674  
Patent No. US20010039335A1  
GENERAL INFORMATION:  
APPLICANT: Jacobs, Kenneth  
APPLICANT: McCoy, John M.  
APPLICANT: Lavallee, Edward R.  
APPLICANT: Collins-Racie, Lisa A.  
APPLICANT: Evans, Cheryl  
APPLICANT: Werberg, David  
APPLICANT: Treacy, Maurice  
APPLICANT: Agostino, Michael J.  
APPLICANT: Steinger II, Robert J.  
APPLICANT: Spaulding, Vikki  
APPLICANT: Wong, Gordon G.  
APPLICANT: Clark, Hilary  
APPLICANT: Fecthel, Kim  
TITLE OF INVENTION: GENETICS INSTITUTE, INC.  
FILE REFERENCE: 6055-64X  
CURRENT APPLICATION NUMBER: US/09/729,674  
CURRENT FILING DATE: 2000-12-04  
PRIOR APPLICATION NUMBER: 09/539,330  
PRIOR FILING DATE: 2000-03-30  
NUMBER OF SEQ ID NOS: 283  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 173  
LENGTH: 1534  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-729-674-173  
  
Alignment Scores:  
Pred. No.: 1,29e-43 Length: 1534  
Score: 390.00 Matches: 83  
Percent Similarity: 61.62% Conservative: 31  
Best Local Similarity: 44.86% Mismatches: 59  
Query Match: 35.91% Indels: 12  
DB: Gaps: 3  
US-09-966-880A-8 (1-198) x US-09-729-674-173 (1-1534)  
  
QY 6 MetAsnAArgAlyGlyPheLeuTyrglnPheLyAsnValAArgTPAlaLySGlyAArgA 25  
DB 639 ATGATTCACCCACATTCATCTTCACTTAACTTAACTTGGGTCAGAGAGCGCAT 698  
QY 26 GluThrTyLeuCyTyValValValysAArgAseAalathSerPheSerLeuasp 45  
DB 699 GAGACTTACCTGTGTATGAGGTGAGGAGCATGACATGACCTGGTCTGCTGAAC 758  
QY 46 -----PheGlyTyrLeuAArgAAsnLyS-----AsnGlyCyS 55  
DB 759 CAGCGCAGGGGCTTCTATGACACCGCTCCACATAAACAGGTTCTTGAAGCGGC 818  
QY 56 HisValGlnLeuLeuPheLeuAArgTyrlleSerAspTPAAspProGlyAArgCyS 75  
DB 819 CATGACAGGCTGTGCTCTGAGCGTGAATCCCTTTGGAAGCTGAGACCTGAGACCGAGAC 878  
QY 76 TyArgValThrTrpPheThrSerTrpSerProCyTyAAspCyValAArgHisValAla 95  
DB 879 TACAGGTTACCTGCTTCACTCCCTGAGCGCCCTGCTTCACTGAGCCCGCAATGCT 938  
QY 96 AspPheLeuAArgGlyAsnProAsnLeuSerLeuAArgIlePheThrAlaArgLeuTyrPhe 115  
DB 939 AATTCATTCAAAAACAAACACGTCGAGCTGTGACATCTTCACTGCGCCGCACTAT--- 995  
QY 116 CysGluAspAArgLyValAgluProGlnGlyLeuAArgLyHisAArgAlaGlyValGln 135  
DB 996 ---GATGATCAAGAGAGATGTCAGAGGGGCTGCCACCTGGCCGAGGCTGGAGCCCAA 1052  
QY 136 IleAlaIleMetThrPheLyAspTyrglnPheTyrcySTPAsnThrPheValGluAsnHis 155



Db 1053 ATTCAATTAATGACATACAGTAAATTAAAGCACTGTGGACACCTTTGTGACCAACG 1112  
Qy 156 GlnuYgthrPheLysAlaTTrpGluGlyLeuHISgluLnsSerValaLysLeuSerArgGln 175  
Db 1113 GGATGTCCCTCCACGCCCTGGAGTGAAGTATGATGACAGACCCAAAGACTGAGTGGAGG 1172  
Qy 176 LeuArgArgIleLeu 180  
Db 1173 CTCGGGGCCATTCTC 1187

RESULT 8  
US-09-925-300-699  
Sequence 699, Application US/09925300  
Patent No. US20020151681A1  
GENERAL INFORMATION:  
APPLICANT: Craig Rosen,  
APPLICANT: Steve Ruben  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: P101  
CURRENT APPLICATION NUMBER: US/09/925,300  
CURRENT FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05988  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 1890  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 699  
LENGTH: 987  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-925-300-699

Alignment Scores:  
Pred. No.: 4,75e-41 Length: 987  
Score: 369.50 Matches: 79  
Percent Similarity: 58.52% Conservative: 24  
Best Local Similarity: 44.89% Mismatches: 64  
Query Match: 34.02% Indels: 9  
Gaps: 4

US-09-966-880a-8 (1-198) x US-09-925-300-699 (1-987)

Qy 11 PheLeuTYrGlnPheLysAlaValArgTrpAlaLysGlyArgArgGlnTrpTYrLeuGys 30  
Db 147 TTCTACTTCACATTTAAACCTTATGGAGCCAAAGCCAAAGTGGACCAAACTGCTGTGC 206  
Qy 31 TyrValValLys-----ArgArgAspSerAlaThrSerPheSerLeuAspPheGlyTyr 48  
Db 207 TTCAACCGTGAAGATTAAGCGCCGCTCACTTGTCTCTCGAAGACG-----GGCGTC 260  
Qy 49 LeuArgAsn-----LysAsnGlyCysHisValGluLeuLeuPheLeuArgTyr 64  
Db 261 TTCCGAAACCAAGTGGATTCGAGCCCATTTGTATGACAGAAAGTCTCTCTCTGG 320  
Qy 65 IleSerAspTrpAspLeuAspProGlyArgCysTYrArgValThrTrpPheThrSerTrp 84  
Db 321 TTCTGCGACGACATCTCTCTCTTACCAAGTACCAAGTCACTGCTGACATCTTGG 380  
Qy 85 SerProCysTYrAspCysAlaArgHisValaLysPheLeuArgGlyAsnProAsnLeu 104  
Db 381 AGCCCTTCCAGACTGTGACAGGGAGGTGCGCAGTTCTCGGCGACAGACAGACGTG 440  
Qy 105 SerLeuArgIlePheThrAlaArgLeuTYrPheCysGluAspArgLysAlaGluProGlu 124  
Db 441 AATTCACCATCTTCACCGCCGCTTACTACTTCAG--TATCATGTTTACAGAG 497  
Qy 125 GlyLeuArgArgLysHisArgAlaGlyValGlnIleAlaIleMetThrPheLysAspTyr 144  
Db 498 GGGCTCCGACCTGAGTCAAGAGAGGGTCTGTGAGATCAAGACATTAAGAAATTT 557  
Qy 145 PheTYrCysTrpAsnThrPheValGluAsnHisgluArgThrPheLysAlaTrpGluGly 164

Db 558 AATATTGTGTGGAAACCTTTGTGTACATGATATATGACCATTCACCTTGAAGGA 617  
Qy 165 LeuHISgluLnsSerValaLysLeuSerArgGlnLeuArgArgIleLeu 180  
Db 618 TTTAAACCAACTTTCGACTTCTGAAAGAGCTACGGAGAGTCTC 665

RESULT 9  
US-10-098-841-268  
Sequence 268, Application US/10098841  
Publication No. US20020197679A1  
GENERAL INFORMATION:  
APPLICANT: Tang, Y. Tom  
APPLICANT: Liu, Chenghua  
APPLICANT: Asundi, Vinod  
APPLICANT: Xu, Chongjun  
APPLICANT: Zhou, Ping  
APPLICANT: Ma, Yungqing  
APPLICANT: Wang, Jian-Rui  
APPLICANT: Zhao, Qing A.  
APPLICANT: Ren, Feiyan  
APPLICANT: Chen, Rui-Hong  
APPLICANT: Wang, Dunrui  
APPLICANT: Wang, Zhimei  
APPLICANT: Wehrman, Tom  
APPLICANT: Zhang, Jie  
APPLICANT: Qian, Xiaohong B.  
TITLE OF INVENTION: No. US20020197679A1e1 Nucleic Acids and  
FILE REFERENCE: 784CIP2  
CURRENT APPLICATION NUMBER: US/10/098,841  
CURRENT FILING DATE: 2002-03-13  
PRIOR APPLICATION NUMBER: 09/598,042  
PRIOR FILING DATE: 2000-06-20  
PRIOR APPLICATION NUMBER: 09/552,317  
PRIOR FILING DATE: 2000-04-25  
PRIOR APPLICATION NUMBER: 09/488,725  
PRIOR FILING DATE: 2000-01-21  
NUMBER OF SEQ ID NOS: 331  
SOFTWARE: pt\_FL\_genes Version 1.0  
SEQ ID NO 268  
LENGTH: 1143  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (128)..(700)  
US-10-098-841-268

Alignment Scores:  
Pred. No.: 5.94e-41 Length: 1143  
Score: 369.50 Matches: 79  
Percent Similarity: 58.52% Conservative: 24  
Best Local Similarity: 44.89% Mismatches: 64  
Query Match: 34.02% Indels: 9  
Gaps: 4

US-09-966-880a-8 (1-198) x US-10-098-841-268 (1-1143)

Qy 11 PheLeuTYrGlnPheLysAlaValArgTrpAlaLysGlyArgArgGlnTrpTYrLeuGys 30  
Db 176 TTCTACTTCACATTTAAACCTTATGGAGCCAAAGTGGACCAAACTGCTGTGC 235  
Qy 31 TyrValValLys-----ArgArgAspSerAlaThrSerPheSerLeuAspPheGlyTyr 48  
Db 236 TTCAACCGTGAAGATTAAGCGCCGCTCACTTGTCTCTCGAAGACG-----GGCGTC 289  
Qy 49 LeuArgAsn-----LysAsnGlyCysHisValGluLeuLeuPheLeuArgTyr 64  
Db 290 TTCCGAAACCAAGTGGATTCGAGCCCATTTGTATGACAGAAAGTGTCTCTCTGG 349  
Qy 65 IleSerAspTrpAspLeuAspProGlyArgCysTYrArgValThrTrpPheThrSerTrp 84



Db 350 TTCTGGACGACATCTGCTCTTAACACAAAGTACAGGTCCAGCTGTACACTCTGG 409  
 QY 85 SerProCyteTyraSPCyAlaArgHisValAlaAspPheLeuArgGlyAsnProAsnLeu 104  
 Db 410 AGCCCTTGGCCAGACCTGTGCGAGGGAGGTGGCCGAGCTTCTGCGCAGACACAGCAAGT 469  
 QY 105 SerLeuArgIlePheThrAlaArgLeuTyPheCySGluAspArgGlyAlaGluProGlu 124  
 Db 470 AATCTACACATCTTCCACCCCGCTCTACTACTCTCCAG---TATCCATGTACACAGAG 526  
 QY 125 GlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThrPheLysAspTy 144  
 Db 527 GGGCTCCGACGCTGTGCTGACGAGAGGGGTGCTGCTGAGAGATGAGCTATGAAAGATT 586  
 QY 145 PheTyrcystrpAsnThrPheValGluAsnHisGluArgThrPheLysAlaTrpGluGly 164  
 Db 587 AATATATTGTGGAAAACTTTGTGTACATGATGATGAGCCATTGACCTTGAGAGGA 646  
 QY 165 LeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180  
 Db 647 TTTAAAAACCACTTTCGACTTCTGAAGAGGCTACGCGAGAGTCTC 694

RESULT 10  
 US-09-880-107-3282  
 ; Sequence 3282, Application US/09880107  
 ; Patent No. US20020142981A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Horne, Darci T.  
 ; APPLICANT: Vockley, Joseph G.  
 ; APPLICANT: Scherf, Dwe  
 ; APPLICANT: Gene Logic, Inc.  
 ; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer  
 ; FILE REFERENCE: 44921-5028-WO  
 ; CURRENT APPLICATION NUMBER: US/09/880,107  
 ; PRIOR FILING DATE: 2001-06-14  
 ; PRIOR APPLICATION NUMBER: US 60/211,379  
 ; PRIOR FILING DATE: 2000-06-14  
 ; PRIOR APPLICATION NUMBER: US 60/237,054  
 ; PRIOR FILING DATE: 2000-10-02  
 ; NUMBER OF SEQ ID NOS: 3950  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 3282  
 ; LENGTH: 1348  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; OTHER INFORMATION: Genbank Accession No. US20020142981A1 U03891  
 ; NAME/KEY: unsure  
 ; LOCATION: (1)-(1348)  
 ; OTHER INFORMATION: n = a or c or g or t  
 US-09-880-107-3282

Alignment Scores:  
 Pred. No.: 5,25e-40 Length: 1348  
 Score: 363.50 Matches: 83  
 Percent Similarity: 58.12% Conservative: 28  
 Best Local Similarity: 43.46% Mismatches: 57  
 Query Match: 33.47% Indels: 23  
 Gaps: 7

US-09-966-880a-8 (1-198) x US-09-880-107-3282 (1-1348)

QY 5 LeuMetAsnArgArgIArgLysPheLeuTyGlnPheLysAsnValArgTrpAlaLysGlyArg 24  
 Db 109 TTGATGATCCACACATATCTCACTTCACTTAACAAT-----GGCATTGGAAG 159  
 QY 25 ArgGlnThrTyLeuCytyrValValLysArgAspSerAlaThrSerPheSerLeu 44  
 Db 160 CATAGACCTTACTGTGCTTACGAGGTGAGCGCTGACCAATGGACCTCGGTCAAGATG 219  
 QY 45 Asp-----PheGlyTyTrLeuArgAsn-----LysAsn-----Gly 54  
 Db 220 GACACAGACAGGGGCTTTTCTACACACAGCGCTAAGATCTTCTGTGCGCTTTACGGC 279

QY 55 CysHisValGluLeuLeuPheLeuArgTyrlIleSerAspTrpAspLeuAspProGlyArg 74  
 Db 280 CGCCATGGAGAGCTGGCGCTTCTTGACCTGTCTTGGAGTGGACCGGCGCCAG 339  
 QY 75 CysTyraGlyValThrTrpPheThrSerTrpSerProCysTyraPhe-----CysAlaArg 92  
 Db 340 ACTTCACAGGTCACCTGTGATCTCTGAGGCCCTGCTTCCCTGGGGCTGTGCCGG 399  
 QY 93 HisValAlaAspPheLeuArgGlyAsnProAsnSerLeuArgIlePheThrAlaArg 112  
 Db 400 GAAATGCTGCTGCTTCTTACAGAAACACACAGTGAATGCGCATCTTCTGCGCCG 459  
 QY 113 LeuTyrcySGluAspArgGlyAlaGluPro-----GluGlyLeuArgArgLeu 129  
 Db 460 ATCTAT-----GATTACGACCCCTATATAGAGAGGCGCTGCAAAATGCTG 504  
 QY 130 HisArgAlaGlyValGlnIleAlaIleMetThrPheLysAspTyrcystrpAsn 149  
 Db 505 CGGATGCTGGGGCCCAAGTCTCCATCATGATGATGATTAAGCATCTGCTGGAC 564  
 QY 150 ThrPheValGluAsnHisGluArgThrPheLysAlaTrpGluGlyLeuHisGluAsnSer 169  
 Db 565 ACCTTTGGACACACAGAGGATGTCCTTCCAGCCCTGGAGTGAATGAGACAGC 624  
 QY 170 ValArgLeuSerArgGlnLeuArgArgIleLeu 180  
 Db 625 CAAGCCCTGAGTGGAGGCTGCGGCGCATTTCTC 657

RESULT 11  
 US-10-247-671-62  
 ; Sequence 62, Application US/10247671  
 ; Publication No. US20030194721A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Mikita, Thomas  
 ; APPLICANT: Shiftman, Dov  
 ; APPLICANT: Porter, Gordon, J.  
 ; APPLICANT: Kaser, Matthew R.  
 ; TITLE OF INVENTION: GENES EXPRESSED IN TREATED FOAM CELLS  
 ; FILE REFERENCE: PA-0050 US  
 ; CURRENT APPLICATION NUMBER: US/10/247,671  
 ; PRIOR FILING DATE: 2002-09-18  
 ; PRIOR APPLICATION NUMBER: 60/323,784  
 ; PRIOR FILING DATE: 2001-09-19  
 ; NUMBER OF SEQ ID NOS: 186  
 ; SOFTWARE: PERL Program  
 ; SEQ ID NO 62  
 ; LENGTH: 823  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc. feature  
 ; OTHER INFORMATION: Incyte ID No. US20030194721A1 135626CB1  
 US-10-247-671-62

Alignment Scores:  
 Pred. No.: 3,58e-38 Length: 823  
 Score: 348.00 Matches: 77  
 Percent Similarity: 59.41% Conservative: 24  
 Best Local Similarity: 45.29% Mismatches: 49  
 Query Match: 32.04% Indels: 20  
 Gaps: 6

US-09-966-880a-8 (1-198) x US-10-247-671-62 (1-823)

QY 26 GluThrTyLeuCytyrValValLysArgArgAspSerAlaThrSerPheSerLeuAsp 45  
 Db 3 AAGACTTACTTGTGCTTACGAGGTGAGCGCTGACCAATGGACCTCGGTCAAGATGAC 62  
 QY 46 -----PheGlyTyTrLeuArgAsn-----LysAsn-----Gly 55  
 Db 63 CAGCACAGGGGCTTTTCTACACACAGCGCTAAGATCTTCTGTGCGCTTTACGGCGC 122



```
OY 56 HisValGluLeuLeuPheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCys 75
DB 123 CATCGAGAGCTGGCGCTTCTTGGACCTGGTTCCTTCTTGGACCTTGGACCCGGCCAGATC 182
OY 76 TyrArgValThrTrpPheThrSerTrpSerProCysTyrAsp-----CysAlaArgHis 93
DB 183 TACAGGGGCACTTGTTATCTCTCGAGGCGCTGCTCTCTCGGGGCTGGCCGGGAAA 242
OY 94 ValAlaAspPheLeuArgGlyAsnProAsnLeuSerLeuArgIlePheThrAlaArgLeu 113
DB 243 GTGGGTGGCTTCCTTCAGAGACACACAGTGAAGCTGGCATCTTCGCTCCCGCATC 302
OY 114 TyrPheCysGluAspArgGlyValGluPro-----GluGlyLeuArgArgLeuHis 130
DB 303 TAT-----GATTATGACCCCTTATATAGAGGCGCTGCAATGCTCCGG 347
OY 131 ArgAlaGlyValGlnIleAlaIleMetThrPheLysAspTyrPheTyrCysTrpAsnThr 150
DB 348 GATCGTGGGGCCCAAGTCCATCATGACCTAGACATAGTGTGATCTGCTGGACACCC 407
OY 151 PheValGluAsnHisGluArgThrPheLysAlaTrpGluGlyLeuHisGluAsnSerVal 170
DB 408 TTGTGTACCGCCAGAGATGTCCTTCAGCCCTGGAGTGAATGACAGAGACACAGCCAA 467
OY 171 ArgLeuSerArgGlnLeuArgArgIleLeu 180
DB 468 GCCCTGAGTGGAGGCTCGGGCCCATTTCTC 497

RESULT 12
US-10-131-827-8890
; Sequence 8890, Application US/10131827
; Publication No. US2004009479A1
; GENERAL INFORMATION:
; APPLICANT: Wohlgenuth, Jay
; APPLICANT: Fry, Kirk
; APPLICANT: Woodward, Robert
; APPLICANT: Ly, Ngoc
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING AND MONITORING AUTOIMMUNE
; TITLE OF INVENTION: CHRONIC INFLAMMATORY DISEASES
; FILE REFERENCE: 506612000120
; CURRENT APPLICATION NUMBER: US/10/131,827
; CURRENT FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: US 10/006,290
; PRIOR FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: US 60/296,764
; PRIOR FILING DATE: 2001-06-08
; NUMBER OF SEQ ID NOS: 9090
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8890
; LENGTH: 1008
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-131-827-8890

Alignment Scores:
Pred. No.: 2,43e-37 Length: 1008
Score: 343.00 Matches: 79
Percent Similarity: 55.50% Conservative: 27
Best Local Similarity: 41.36% Mismatches: 65
Query Match: 31.58% Indels: 20
DB: 15 Gaps: 6

US-09-966-880A-8 (1-198) x US-10-131-827-8890 (1-1008)
OY 5 LeuMetAsnArgArgGlySerPheLeuTyrGlnPheLysAsnValArgTrpAlaLysGlyArg 24
DB 77 CTGATGATCCAGACACATTCATTCACTTTAATAAGACCTTGTGCTCTCGACGG 136
OY 25 ArgGluThrTyrLeuArgTyrValValLysArgArgAspSerAlaThrSerPheSerLeu 44
DB 137 CGCAGACCTACTGTTGCTATGAGTGGAGCGCTGGACATAGGCTGATGCTGATG 196
OY 45 Asp-----PheGlyTyrLeuArgAsn-----LysAsn-----Gly 54
```

```
DB 197 GACCAGCACATGGCGCTTCTATGCAACGAGCTAAGAAATCTTCTGCTTTTACGGC 256
OY 55 CysHisValGluLeuLeuPheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArg 74
DB 257 CCGCATGCGGAGCTGGCGCTTCTTGAACCTGGTTCCTTCTTGGACCTTGGACCCGGCCAG 316
OY 75 CysTyrArgValThrTrpPheThrSerTrpSerProCysTyrAsp-----CysAlaArg 92
DB 317 ATCTACAGGGTCACTTGTTATCTCTCGAGGCGCTGCTCTCTCGGGGCTGGCCGG 376
OY 93 HisValAlaAspPheLeuArgGlyAsnProAsnLeuSerLeuArgIlePheThrAlaArg 112
DB 377 GAAGTCGGCTTCCTTCAGAGAACACACAGTGAAGCTGGCATCTTCGCTCCCGCC 436
OY 113 LeuTyrPheCysGluAspArgGlyValGluPro-----GluGlyLeuArgArgLeu 129
DB 437 ATCTAT-----GATTATGACCCCTTATATAGAGGCGCTGCAATGCTG 481
OY 130 HisArgAlaGlyValGlnIleAlaIleMetThrPheLysAspTyrPheTyrCysTrpAsn 149
DB 482 CCGAGATGCTGGGCGCCAGTCTTCATGACCTTACATGATGATGATGATGATGATGATG 541
OY 150 ThrPheValGluAsnHisGluArgThrPheLysAlaTrpGluGlyLeuHisGluAsnSer 169
DB 542 ACCTTTGTGTACCGCCAGAGATGTCCTTCAGCCCTGGAGTGAATGACAGAGACACAGC 601
OY 170 ValArgLeuSerArgGlnLeuArgArgIleLeu 180
DB 602 CAAGCCCTGAGTGGAGGCTCGGGCCCATTTCTC 634

RESULT 13
US-09-822-830A-359
; Sequence 359, Application US/09822830A
; Patent No. US20020142952A1
; GENERAL INFORMATION:
; APPLICANT: Genetics Institute, Inc.
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Agostino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Galukota, Kamalakar
; APPLICANT: Graham, James R.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6402
; CURRENT APPLICATION NUMBER: US/09/822,830A
; CURRENT FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195,604
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 631
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 359
; LENGTH: 2151
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-830A-359

Alignment Scores:
Pred. No.: 5.02e-32 Length: 2151
Score: 308.50 Matches: 81
Percent Similarity: 58.20% Conservative: 29
Best Local Similarity: 42.86% Mismatches: 60
Query Match: 28.41% Indels: 22
DB: 9 Gaps: 4

US-09-966-880A-8 (1-198) x US-09-822-830A-359 (1-2151)
OY 6 MetAsnArgArgGlySerPheLeuTyrGlnPheLysAsnValArgTrpAlaLysGlyArg 25
DB 1278 ATGATACACCCACATTCATTCACTTTAACAATGAACCTTGGGTGACAGA-CGGCAT 1336
```







; LENGTH: 148  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-966-880A-12

Alignment Scores:  
Pred. No.: 4.81e-27 Length: 148  
Score: 260.00 Matches: 49  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 23.94% Indels: 0  
DB: 9 Gaps: 0

US-09-966-880A-8 (1-198) x US-09-966-880A-12 (1-148)

QY	4	LeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTIPAlaLysGly	23
DB	2	CTCTTGATGAACCGAGGAGAGTTTCTTACCAATTCAAAAATGCCGCTGGGCTAAGGGT	61
QY	24	ArgArgGluThrTyrLeuGlyTyrValIleLysArgArgAspSerAlaThrSerPheSer	43
DB	62	CGGCGTGAAGACCTACCTGTGCTACGTAGAGAGGCGTGAACAGTGTACATCCTTTCA	121
QY	44	LeuAspPheGlyTyrLeuArgAsnLys	52
DB	122	CTGGACTTTGGTTATCTTCGCATATAG	148

Search completed: March 13, 2004, 15:47:23  
Job time : 323 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OK nucleic - nucleic search, using sw model

Run on: March 11, 2004, 17:19:44 ; Search time 44.5137 Seconds  
(without alignments)  
3740.089 Million cell updates/sec

Title: US-09-966-880A-9\_COPY\_1\_300

Perfect score: 300

Sequence: 1 acagacgaatcatcgtccca.....agcctgttccaatccag 300

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued\_Patents\_NA.\*  
2: /cgn2\_6/pdata/2/ina/5A.COMB.seq:\*  
3: /cgn2\_6/pdata/2/ina/5B.COMB.seq:\*  
4: /cgn2\_6/pdata/2/ina/5A.COMB.seq:\*  
5: /cgn2\_6/pdata/2/ina/5B.COMB.seq:\*  
6: /cgn2\_6/pdata/2/ina/Backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	34.6	11.5 128779	4 US-09-497-855A-38	Sequence 38, Appl
2	33.6	11.2 795	4 US-09-107-532A-3159	Sequence 3159, Ap
3	33	11.0 63000	4 US-09-780-172-18	Sequence 18, Appl
4	32.4	10.8 12619	4 US-09-616-289-49	Sequence 49, Appl
5	32.2	10.7 2413	4 US-09-399-913-48	Sequence 48, Appl
6	32.2	10.7 7218	1 US-08-232-463-14	Sequence 14, Appl
7	32.2	10.7 19233	4 US-10-204-708-46	Sequence 46, Appl
8	31.8	10.6 890	4 US-09-634-238-62	Sequence 62, Appl
9	31.2	10.4 1063	4 US-09-838-586-3	Sequence 3, Appl
10	31.2	10.4 1063	4 US-09-838-586-3	Sequence 3, Appl
11	31	10.3 786431	4 US-09-751-389-3	Sequence 3, Appl
12	30.8	10.3 854	4 US-09-247-155-67	Sequence 67, Appl
13	30.8	10.3 90541	4 US-09-759-359A-3	Sequence 3, Appl
14	30.6	10.2 460	4 US-09-621-976-11559	Sequence 11559, A
15	30.2	10.1 640681	4 US-09-790-988-1	Sequence 1, Appl
16	30	10.0 2265	3 US-09-369-618-3	Sequence 3, Appl
17	30	10.0 2265	3 US-09-369-617-3	Sequence 3, Appl
18	30	10.0 1664976	4 US-08-916-421B-1	Sequence 1, Appl
19	29.8	9.9 31491	3 US-09-360-186-1	Sequence 1, Appl
20	29.8	9.9 50000	4 US-09-146-053-4	Sequence 4, Appl
21	29.8	9.9 1664976	4 US-08-916-421B-1	Sequence 1, Appl
22	29.6	9.9 358	4 US-09-621-976-8138	Sequence 8138, Ap
23	29.6	9.9 40328	4 US-08-742-185-102	Sequence 102, App
24	29.6	9.9 43795	3 US-08-742-185-101	Sequence 101, App
25	29.6	9.9 202001	4 US-09-734-674-3	Sequence 3, Appl
26	29.4	9.8 2218	4 US-09-673-395A-547	Sequence 547, App
27	29.4	9.8 2882	1 US-08-393-219-2	Sequence 2, Appl

28	29.4	9.8 5053	4 US-09-620-312D-590	Sequence 590, App
29	29.4	9.8 5427	3 US-09-009-913-2	Sequence 2, Appl
30	29.4	9.8 5510	3 US-09-009-913-3	Sequence 3, Appl
31	29.4	9.8 5667	3 US-09-009-913-4	Sequence 4, Appl
32	29.4	9.8 392000	4 US-10-027-983-11	Sequence 11, Appl
33	29.2	9.7 1422	4 US-09-485-077A-15	Sequence 15, Appl
34	29.2	9.7 1424	4 US-09-485-076-1	Sequence 1, Appl
35	29.2	9.7 1493	6 5340934-5	Patent No. 5340934
36	29.2	9.7 2116	4 US-09-377-497-11	Sequence 11, Appl
37	29.2	9.7 2575	1 US-08-135-511-33	Sequence 33, Appl
38	29.2	9.7 2575	1 US-08-483-852-10	Sequence 10, Appl
39	29.2	9.7 2575	1 US-08-361-458-5	Sequence 5, Appl
40	29.2	9.7 2575	1 US-08-477-953-10	Sequence 10, Appl
41	29.2	9.7 2575	1 US-08-187-453-33	Sequence 33, Appl
42	29.2	9.7 2575	2 US-08-477-952-10	Sequence 10, Appl
43	29.2	9.7 4646	4 US-09-377-497-67	Sequence 67, Appl
44	29.2	9.7 5152	4 US-10-204-708-74	Sequence 74, Appl
45	29.2	9.7 580073	4 US-08-545-528D-1	Sequence 1, Appl

#### ALIGNMENTS

RESULT 1  
US-09-497-855A-38  
Sequence 38, Application US/09497855A  
Patent No. 6605432  
GENERAL INFORMATION:  
APPLICANT: Huang, Tim  
TITLE OF INVENTION: HIGH-THROUGHPUT METHODS FOR DETECTING DNA METHYLATION  
FILE REFERENCE: UMO1523  
CURRENT APPLICATION NUMBER: US/09/497,855A  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: 60/120,592  
PRIOR FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: 60/118,760  
PRIOR FILING DATE: 1999-02-05  
NUMBER OF SEQ ID NOS: 54  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 38  
LENGTH: 128779  
TYPE: DNA  
ORGANISM: Homo sapiens;  
US-09-497-855A-38

Query Match 11.5%; Score 34.6; DB 4; Length 128779;  
Best Local Similarity 67.1%; Pred. No. 2.4;  
Matches 49; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 68 AGGCTTGAGGACAGAGAGAGACCTTACCTGATGCTTACGATTCGATCCCTAG 127  
DB 53956 AGACTGTAACCTTGAGAGAGTGGGACCATCTGCTTATCAGTATCCCTGG 54015

QY 128 CACCTGGCATAGT 140  
DB 54016 TACCTGGCAGAGT 54028

RESULT 2  
US-09-107-532A-3159  
Sequence 3159, Application US/09107532A  
Patent No. 6583275  
GENERAL INFORMATION:  
APPLICANT: Lynn A Doucette-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 7310  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA



ZIP: 02354  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD/ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,532A  
FILING DATE: 30-Jun-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051,571  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Arinello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: CTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 3159:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 795 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHEICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...795  
SEQUENCE DESCRIPTION: SEQ ID NO: 3159:  
US-09-107-532A-3159

Query Match  
Best Local Similarity 11.2%; Score 33.6; DB 4; Length 795;  
Matches 63; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

QY 189 GCATATGGGCTACAGTACGAGAGAGAAATTAATCTTTATAGACCAAGTTATAGAG 248  
DB 397 GCAACTCACTAAAGAGAGAGAAAGTATGAGAGAGAGAGAGAGAGAGAGAGAG 456  
QY 249 CACAAATTAAGTCTTTATTTGAAGATCTTACCTGCTGTTTCCAAATTCAG 300  
DB 457 TAGGAGCTAAATCTCTTTATTTAGATAGCATTTATCTTTACAGATTTCAG 508

RESULT 3  
US-09-780-172-18/c  
Sequence 18, Application US/09780172  
Patent No. 6607916  
GENERAL INFORMATION:  
APPLICANT: Robert McKay  
APPLICANT: Susan M. Freiler  
TITLE OF INVENTION: ANTISENSE MODULATION OF CASEIN KINASE 2-ALPHA EXPRESSION  
FILE REFERENCE: PUS-0159  
CURRENT APPLICATION NUMBER: US/09/780,172  
FILING DATE: 2001-02-08  
NUMBER OF SEQ ID NOS: 96  
SEQ ID NO 18  
LENGTH: 63000  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
US-09-780-172-18

Query Match  
Best Local Similarity 11.0%; Score 33; DB 4; Length 63000;  
Matches 58.8%; Pred. No. 5.6;

Matches 57; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 136 ATAGTTCCATTACAGTACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 195  
DB 24931 ATAGTGCAGAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 24872  
QY 196 GGTACAGTACGAG 232  
DB 24871 CAACCGAG 24835

RESULT 4  
US-09-616-288-49  
Sequence 49, Application US/09616289  
Patent No. 6632923  
GENERAL INFORMATION:  
APPLICANT: Lees, Robert S.  
APPLICANT: Lees, Robert S.  
APPLICANT: Law, Simon W.  
APPLICANT: Arjona, Anibal A.  
TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING  
TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING  
FILE REFERENCE: 10797-004001  
CURRENT APPLICATION NUMBER: US/09/616,289  
FILING DATE: 2000-07-14  
PRIOR APPLICATION NUMBER: US 09/517,849  
PRIOR FILING DATE: 2000-03-02  
PRIOR APPLICATION NUMBER: US 08/979,608  
PRIOR FILING DATE: 1997-11-26  
PRIOR APPLICATION NUMBER: US 60/031,930  
PRIOR FILING DATE: 1996-11-27  
PRIOR APPLICATION NUMBER: US 60/048,547  
PRIOR FILING DATE: 1997-06-03  
NUMBER OF SEQ ID NOS: 53  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 49  
LENGTH: 12619  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-616-288-49

Query Match  
Best Local Similarity 10.8%; Score 32.4; DB 4; Length 12619;  
Matches 90; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 106 TTGCTAGAGTATGATCCCTAGCAGCTGAGTATGCTTATTAAGTACAGTGAAGTA 165  
DB 9961 TTGTCACCTTTGATATCCGAGCAGTACAGTACAGTACAGTACAGTATTCATA 10020  
QY 166 TCTACTCAGTAAATATGAGATGATATGAGTACAGTACAGTACAGTACAGTAC 225  
DB 10021 AATATTTTGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 10080  
QY 226 TAATAGCAGTATGAG 285  
DB 10081 TATTTTCAATTAAGTACAGTACAGTACAGTACAGTACAGTACAGTACAGTAC 10140  
QY 286 GTTTC 291  
DB 10141 TCTTTC 10146

RESULT 5  
US-09-399-913-48  
Sequence 48, Application US/09399913  
Patent No. 6361971  
GENERAL INFORMATION:  
APPLICANT: Rhodes, Kenneth  
APPLICANT: Betty, Maria  
APPLICANT: Ling, Hwai-Ping  
APPLICANT: An, Wenglian  
TITLE OF INVENTION: POTASSIUM CHANNEL INTERACTORS AND USES THEREFOR



CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/232,463  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/07/935,313  
 FILING DATE:  
 APPLICATION NUMBER: EP 91 114 300.6  
 FILING DATE: 26-AUG-1991  
 ATTORNEY/AGENT INFORMATION:  
 NAME: BENT, Stephen A.  
 REGISTRATION NUMBER: 29,768  
 REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (703) 936-9300  
 TELEFAX: (703) 683-4109  
 TELEX: 899149  
 INFORMATION FOR SEQ ID NO: 14:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 7218 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 IMMEDIATE SOURCE:  
 CLONE: PTZgpt-F18  
 US-08-232-463-14

Query Match	10.7%	Score 32.2;	DB 1;	Length 7218;
Best Local Similarity	3.8%;	Pred. No. 4.3;		
Matches	10;	Conservative 144;	Mismatches 107;	Indels 0;
				Gaps 0.

QY 1 AAGAGCAATACATGCTCCAAAGTAGGGCTATTGATTGAAAATCATCAAGGTATAGATG 60  
Db 1453 AAGATGAAGAATTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1394  
QY 61 GATATCAAGGCTGTGGCAGGAAGAGCAGAGCCCTAGCTCATCTTGCTTGACATTGCA 120  
Db 1393 RRR 1333  
QY 121 TCCCTAGCACTGTGCATAGTTTCATTAAACAGTAGGCGATGAATATCTACAGTGATA 180  
Db 1333 RRR 1274  
QY 181 AATGAGATGCAATATGGGCTACAGTAGAGAGAGAAAATAATTTAATAGACCAAGTTC 240  
Db 1273 RRR 1214  
QY 241 TATGAGACCAAAATTAAG 261  
Db 1213 RRR 1193

RESULT 7  
US-10-204-708-46/C  
; Sequence 46, Application US/10204708  
; Patent No. 6677731.  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENBROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication  
; TITLE OF INVENTION: by Assessing DNA Methylation  
; FILE REFERENCE: 5013.1012  
; CURRENT APPLICATION NUMBER: US/10/204.708  
; CURRENT FILING DATE: 2003-05-06  
; PRIOR APPLICATION NUMBER: PCT/EP01/03971  
; PRIOR FILING DATE: 2001-04-06  
; PRIOR APPLICATION NUMBER: DE 10019058.8  
; PRIOR FILING DATE: 2000-04-06  
; PRIOR APPLICATION NUMBER: DE 10019173.8  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: DE 10032529.7  
; PRIOR FILING DATE: 2000-06-30















GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 11, 2004, 17:19:49 ; Search time 526.309 Seconds  
(without alignments)  
2098.537 Million cell updates/sec

Title: US-09-966-880A-9\_COPY\_1\_300

Perfect score: 300  
Sequence: 1 acagacgaatacatgctca.....agcctgttcccaattcag 300

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 2432557 seqs, 1840798884 residues

Total number of hits satisfying chosen parameters: 4865114

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications NA.\*  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
17: /cgn2\_6/ptodata/2/pubpna/US6\_NEW\_PUB.seq:\*  
18: /cgn2\_6/ptodata/2/pubpna/US6\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being pinned, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	300	100.0	5514 9 US-09-966-880A-9	Sequence 9, Appl1
2	277.2	92.4	461 15 US-10-027-632-297069	Sequence 297069, Ap
3	277.2	92.4	461 15 US-10-027-632-297070	Sequence 297070, Ap
4	277.2	92.4	461 15 US-10-027-632-297071	Sequence 297071, Ap
5	273	91.0	810 15 US-10-027-632-4416	Sequence 4416, Ap
6	273	91.0	810 15 US-10-027-632-4417	Sequence 4417, Ap
7	273	91.0	810 15 US-10-027-632-4418	Sequence 4418, Ap
8	40.2	13.4	1210 15 US-10-027-632-214410	Sequence 214410, Ap
9	39.6	13.2	737 15 US-10-027-632-8782	Sequence 8782, Ap
10	39.6	13.2	11216 10 US-09-764-872-490	Sequence 490, Appl1
11	37.6	12.5	600 14 US-10-312-841-2	Sequence 841, Ap
12	37.4	12.4	470 15 US-10-029-386-5897	Sequence 5897, Ap
13	37.2	12.4	470 15 US-10-027-632-270562	Sequence 270562, Ap
14	37.2	12.4	8943 16 US-10-257-166-47	Sequence 47, Appl1
15	36.2	12.1	637 15 US-10-027-632-247353	Sequence 247353, Ap

16	36	12.0	6242 14 US-10-311-455-2121	Sequence 2121, Ap
17	35.4	11.8	580 15 US-10-027-632-32647	Sequence 32647, A
18	35.2	11.7	565 15 US-10-027-632-190320	Sequence 190320, A
19	35.2	11.7	585 15 US-10-027-632-190321	Sequence 190321, A
20	35.2	11.7	1826 15 US-10-027-632-43119	Sequence 43119, A
21	35.2	11.7	1826 15 US-10-027-632-43120	Sequence 43120, A
22	35.2	11.7	3529 15 US-10-369-493-46354	Sequence 46354, A
23	35	11.7	366 9 US-09-960-352-1681	Sequence 1681, Ap
24	35	11.7	379 9 US-09-960-352-6621	Sequence 6621, Ap
25	35	11.7	389 9 US-09-960-352-14093	Sequence 14093, A
26	35	11.7	413 9 US-09-960-352-2490	Sequence 2490, Ap
27	35	11.7	414 9 US-09-960-352-2489	Sequence 2489, Ap
28	35	11.7	414 9 US-09-960-352-10772	Sequence 10772, A
29	35	11.7	416 9 US-09-960-352-8509	Sequence 8509, Ap
30	35	11.7	426 9 US-09-960-352-6031	Sequence 6031, Ap
31	34.8	11.6	607 14 US-10-196-846-9535	Sequence 9535, Ap
32	34.6	11.5	2000 9 US-09-938-842A-4728	Sequence 4728, Ap
33	34.6	11.5	2000 11 US-09-938-842A-4728	Sequence 4728, Ap
34	34.6	11.5	128779 14 US-10-081-327-38	Sequence 38, Appl1
35	34.2	11.4	360 12 US-10-424-599-99764	Sequence 99764, A
36	34.2	11.4	392 9 US-09-960-352-3298	Sequence 3298, Ap
37	34.2	11.4	634 15 US-10-027-632-200698	Sequence 200698, Ap
38	34	11.3	403 9 US-09-960-352-275	Sequence 275, Ap
39	34	11.3	695 15 US-10-027-632-27031	Sequence 27031, A
40	34	11.3	1060 14 US-10-195-730-24	Sequence 24, Appl1
41	34	11.3	3133 15 US-10-027-632-25969	Sequence 25969, Ap
42	34	11.3	3133 15 US-10-027-632-259970	Sequence 259970, Ap
43	34	11.3	3133 15 US-10-027-632-259971	Sequence 259971, Ap
44	34	11.3	3133 15 US-10-027-632-259972	Sequence 259972, Ap
45	34	11.3	3252 15 US-10-027-632-113786	Sequence 113786, Ap

#### ALIGNMENTS

RESULT 1  
US-09-966-880A-9  
Sequence 9, Application US/09966880A  
Patent No. US20020164743A1  
GENERAL INFORMATION:  
APPLICANT: Honjo, Tetsuki  
APPLICANT: Muramatsu, Masamichi  
TITLE OF INVENTION: NOVEL CITIDINE DEAMINASE  
FILE REFERENCE: 06501-088001  
CURRENT APPLICATION NUMBER: US/09/966, 880A  
CURRENT FILING DATE: 2001-09-28  
PRIOR APPLICATION NUMBER: PCT/JP00/01918  
PRIOR FILING DATE: 2000-03-28  
PRIOR APPLICATION NUMBER: JP 11-371382  
PRIOR FILING DATE: 1999-12-27  
PRIOR APPLICATION NUMBER: JP 11-178939  
PRIOR FILING DATE: 1999-06-24  
PRIOR APPLICATION NUMBER: JP 11-87192  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 9  
LENGTH: 5514  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: intron  
LOCATION: (1)...(1031)  
FEATURE:  
NAME/KEY: exon  
LOCATION: (1032)...(1118)  
FEATURE:  
NAME/KEY: intron  
LOCATION: (1119)...(5514)  
US-09-966-880A-9  
Query Match 100.0%; Score 300, DB 9, Length 5514;  
Best Local Similarity 100.0%; Pred. No. 9, 1e-79;



Matches 300; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAGAGCAATAGATGGTCCAGCTAGGCTATTGATTTGAAATCATCAAGATATAGT 60  
DB 1 ACAGAGCAATAGATGGTCCAGCTAGGCTATTGATTTGAAATCATCAAGATATAGT 60

QY 61 GATCAAGGCTTGAGGCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120  
DB 61 GATCAAGGCTTGAGGCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120

QY 121 TCCCTGACCTGGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAG 180  
DB 121 TCCCTGACCTGGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAG 180

QY 181 AATAGATGATATGGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAG 240  
DB 181 AATAGATGATATGGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAG 240

QY 241 TATGAGAGCAAAATTAAGTCTTTTATTTGAGATCTTGAAGCTTTTCCAAATTCAG 300  
DB 241 TATGAGAGCAAAATTAAGTCTTTTATTTGAGATCTTGAAGCTTTTCCAAATTCAG 300

RESULT 2  
US-10-027-632-297069  
; Sequence 297069, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT FILING DATE: US/10/027,632  
; PRIOR FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/118,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 297069  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-297069

Query Match 92.4%; Score 277.2; DB 15; Length 461;  
Best Local Similarity 99.3%; Pred. No. 1.7e-72;  
Matches 276; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

DB 363 AATAGATGATATGGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAG 422  
QY 241 TATGAGAGCAAAATTAAGTCTTTTATTTGAGATCTTGAAGCTTTTCCAAATTCAG 278  
DB 423 TATGAGAGCAAAATTAAGTCTTTTATTTGAGATCTTGAAGCTTTTCCAAATTCAG 460

RESULT 3  
US-10-027-632-297070  
; Sequence 297070, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT FILING DATE: US/10/027,632  
; PRIOR FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 297070  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-297070

Query Match 92.4%; Score 277.2; DB 15; Length 461;  
Best Local Similarity 99.3%; Pred. No. 1.7e-72;  
Matches 276; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAGAGCAATAGATGGTCCAGCTAGGCTATTGATTTGAAATCATCAAGATATAGT 60  
DB 183 ACAGAGCAATAGATGGTCCAGCTAGGCTATTGATTTGAAATCATCAAGATATAGT 242

QY 61 GATCAAGGCTTGAGGCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120  
DB 243 GATCAAGGCTTGAGGCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 302

QY 121 TCCCTGACCTGGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAG 180  
DB 303 TCCCTGACCTGGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAG 352

QY 181 AATAGATGATATGGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAG 240  
DB 363 AATAGATGATATGGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAGGCTAG 422

QY 241 TATGAGAGCAAAATTAAGTCTTTTATTTGAGATCTTGAAGCTTTTCCAAATTCAG 278  
DB 423 TATGAGAGCAAAATTAAGTCTTTTATTTGAGATCTTGAAGCTTTTCCAAATTCAG 460

RESULT 4  
US-10-027-632-297071  
; Sequence 297071, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide



;; TITLE OF INVENTION: Polymorphisms in the Human Genome  
;; FILE REFERENCE: 108827.129  
;; CURRENT APPLICATION NUMBER: US/10/027,632  
;; CURRENT FILING DATE: 2002-04-30  
;; PRIOR APPLICATION NUMBER: US 60/218,006  
;; PRIOR FILING DATE: 2000-07-12  
;; PRIOR APPLICATION NUMBER: US 60/198,676  
;; PRIOR FILING DATE: 2000-04-20  
;; PRIOR APPLICATION NUMBER: US 60/193,483  
;; PRIOR FILING DATE: 2000-03-29  
;; PRIOR APPLICATION NUMBER: US 60/185,218  
;; PRIOR FILING DATE: 2000-02-24  
;; PRIOR APPLICATION NUMBER: US 60/167,363  
;; PRIOR FILING DATE: 1999-11-23  
;; PRIOR APPLICATION NUMBER: US 60/156,358  
;; PRIOR FILING DATE: 1999-09-28  
;; PRIOR APPLICATION NUMBER: US 60/146,002  
;; PRIOR FILING DATE: 1999-08-09  
;; NUMBER OF SEQ ID NOS: 325720  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 297071  
;; LENGTH: 461  
;; TYPE: DNA  
;; ORGANISM: Human  
US-10-027-632-297071

Query Match 92.4%; Score 277.2; DB 15; Length 461;  
Best Local Similarity 99.3%; Pred. No. 1.7e-72;  
Matches 276; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAGCGAATACATGTCTCCAGCTAGGGCTATTGATTGAAATCATCAAGGTATAGTG 60  
DB 183 ACAGCGAATACATGTCTCCAGCTAGGGCTATTGATTGAAATCATCAAGGTATAGTG 242  
QY 61 GATCAAAAGGCTTGAGGCGAAGAGAGAGACCTAGCTGCAATGCTTAGCATTTCA 120  
DB 243 GATCAAAAGGCTTGAGGCGAAGAGAGAGACCTAGCTGCAATGCTTAGCATTTCA 302  
QY 121 TCCCTAGCACTGGCATGTTTCCATTAACTAGTGCATGAAGTACTACTCAAGTGAATA 180  
DB 303 TCCCTAGCACTGGCATGTTTCCATTAACTAGTGCATGAAGTACTACTCAAGTGAATA 362  
QY 161 AATAGAAATGCATATGGGCTACAGTACGAGAGAGAAATTAATCTTTAATAGCAAGTTC 240  
DB 363 AATAGAAATGCATATGGGCTACAGTACGAGAGAGAAATTAATCTTTAATAGCAAGTTC 422  
QY 241 TATGAGAGCAAAATTAAGCTTTTATTGAAGATC 278  
DB 423 TATGAGAGCAAAATTAAGCTTTTATTGAAGATC 460

RESULT 5  
US-10-027-632-4416

;; Sequence 4416, Application US/10027632  
;; Publication No. US20030204075A9  
;; GENERAL INFORMATION:  
;; APPLICANT: Wang, David G.  
;; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
;; FILE REFERENCE: 108827.129  
;; CURRENT APPLICATION NUMBER: US/10/027,632  
;; CURRENT FILING DATE: 2002-04-30  
;; PRIOR APPLICATION NUMBER: US 60/218,006  
;; PRIOR FILING DATE: 2000-07-12  
;; PRIOR APPLICATION NUMBER: US 60/198,676  
;; PRIOR FILING DATE: 2000-04-20  
;; PRIOR APPLICATION NUMBER: US 60/193,483  
;; PRIOR FILING DATE: 2000-03-29  
;; PRIOR APPLICATION NUMBER: US 60/185,218  
;; PRIOR FILING DATE: 2000-02-24  
;; PRIOR APPLICATION NUMBER: US 60/167,363  
;; PRIOR FILING DATE: 1999-11-23  
;; PRIOR APPLICATION NUMBER: US 60/156,358

;; PRIOR FILING DATE: 1999-09-28  
;; PRIOR APPLICATION NUMBER: US 60/146,002  
;; PRIOR FILING DATE: 1999-08-09  
;; NUMBER OF SEQ ID NOS: 325720  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 4416  
;; LENGTH: 810  
;; TYPE: DNA  
;; ORGANISM: Human  
US-10-027-632-4416

Query Match 91.0%; Score 273; DB 15; Length 810;  
Best Local Similarity 99.3%; Pred. No. 4e-71;  
Matches 273; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACAGCAATACATGTCTCCAGCTAGGGCTATTGATTGAAATCATCAAGGTATAGTG 60  
DB 536 ACAGCAATACATGTCTCCAGCTAGGGCTATTGATTGAAATCATCAAGGTATAGTG 595  
QY 61 GATCAAAAGGCTTGAGGCGAAGAGAGAGACCTAGCTGCAATGCTTAGCATTTCA 120  
DB 596 GATCAAAAGGCTTGAGGCGAAGAGAGAGACCTAGCTGCAATGCTTAGCATTTCA 655  
QY 121 TCCCTAGCACTGGCATGTTTCCATTAACTAGTGCATGAAGTACTACTCAAGTGAATA 180  
DB 656 TCCCTAGCACTGGCATGTTTCCATTAACTAGTGCATGAAGTACTACTCAAGTGAATA 715  
QY 181 AATAGAAATGCATATGGGCTACAGTACGAGAGAGAAATTAATCTTTAATAGCAAGTTC 240  
DB 716 AATAGAAATGCATATGGGCTACAGTACGAGAGAGAAATTAATCTTTAATAGCAAGTTC 775  
QY 241 TATGAGAGCAAAATTAAGCTTTTATTGAAG 275  
DB 776 TATGAGAGCAAAATTAAGCTTTTATTGAAG 810

RESULT 6

US-10-027-632-4417

;; Sequence 4417, Application US/10027632  
;; Publication No. US20030204075A9  
;; GENERAL INFORMATION:  
;; APPLICANT: Wang, David G.  
;; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
;; FILE REFERENCE: 108827.129  
;; CURRENT APPLICATION NUMBER: US/10/027,632  
;; CURRENT FILING DATE: 2002-04-30  
;; PRIOR APPLICATION NUMBER: US 60/218,006  
;; PRIOR FILING DATE: 2000-07-12  
;; PRIOR APPLICATION NUMBER: US 60/198,676  
;; PRIOR FILING DATE: 2000-04-20  
;; PRIOR APPLICATION NUMBER: US 60/193,483  
;; PRIOR FILING DATE: 2000-03-29  
;; PRIOR APPLICATION NUMBER: US 60/185,218  
;; PRIOR FILING DATE: 2000-02-24  
;; PRIOR APPLICATION NUMBER: US 60/167,363  
;; PRIOR FILING DATE: 1999-11-23  
;; PRIOR APPLICATION NUMBER: US 60/156,358  
;; PRIOR FILING DATE: 1999-09-28  
;; PRIOR APPLICATION NUMBER: US 60/146,002  
;; PRIOR FILING DATE: 1999-08-09  
;; NUMBER OF SEQ ID NOS: 325720  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 4417  
;; LENGTH: 810  
;; TYPE: DNA  
;; ORGANISM: Human  
US-10-027-632-4417

Query Match 91.0%; Score 273; DB 15; Length 810;  
Best Local Similarity 99.3%; Pred. No. 4e-71;  
Matches 273; Conservative 1; Mismatches 1; Indels 0; Gaps 0;



QY 1 ACAGACCAATTCATGCTCCAGCTAGGGCTATTGATTTGAAATTCATCAAGGTATGAG 60  
DB 536 ACAGACCAATTCATGCTCCAGCTAGGGCTATTGATTTGAAATTCATCAAGGTATGAG 595  
QY 61 GTATCAAAAGGCTTAGGAG 120  
DB 596 GTATCAAAAGGCTTAGGAG 655  
QY 121 TCCCTAGACCTGGGATGATTTCCATTACAGTAGAGATGAGATCTACTAGTGAATA 180  
DB 656 TCCCTAGACCTGGGATGATTTCCATTACAGTAGAGATGAGATCTACTAGTGAATA 715  
QY 181 AATGAAATGCAATGAGGCTACAGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
DB 716 AATGAAATGCAATGAGGCTACAGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 775  
QY 241 TATGAGACCAAAATTAAGTCTTTTATTGAG 275  
DB 776 TATGAGACCAAAATTAAGTCTTTTATTGAG 810

## RESULT 7

US-10-027-632-4418  
; Sequence 4418, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4418  
; LENGTH: 810  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-4418

Query Match 91.0%; Score 273; DB 15; Length 810;  
Best Local Similarity 99.3%; Pred. No. 4e-71;  
Matches 273; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACAGACCAATTCATGCTCCAGCTAGGGCTATTGATTTGAAATTCATCAAGGTATGAG 60  
DB 536 ACAGACCAATTCATGCTCCAGCTAGGGCTATTGATTTGAAATTCATCAAGGTATGAG 595  
QY 61 GTATCAAAAGGCTTAGGAG 120  
DB 596 GTATCAAAAGGCTTAGGAG 655  
QY 121 TCCCTAGACCTGGGATGATTTCCATTACAGTAGAGATGAGATCTACTAGTGAATA 180  
DB 656 TCCCTAGACCTGGGATGATTTCCATTACAGTAGAGATGAGATCTACTAGTGAATA 715  
QY 181 AATGAAATGCAATGAGGCTACAGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240  
DB 716 AATGAAATGCAATGAGGCTACAGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 775

QY 241 TATGAGACCAAAATTAAGTCTTTTATTGAG 275  
DB 776 TATGAGACCAAAATTAAGTCTTTTATTGAG 810

## RESULT 8

US-10-027-632-214410  
; Sequence 214410, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 214410  
; LENGTH: 1210  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-214410

Query Match 13.4%; Score 40.2; DB 15; Length 1210;  
Best Local Similarity 62.4%; Pred. No. 0.16;  
Matches 63; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 15 GTTCCAAAGCTAGGGCTATTGATTTGAAATTCATCAAGGTATGAGGCTTG 74  
DB 828 GGGAGAAAGCTAATGATACAAAGGTGAAATTCATCAATGATGATGATGATGATGAT 887  
QY 75 AGGAG 115  
DB 888 GGGAG 928

## RESULT 9

US-10-027-632-8782/c  
; Sequence 8782, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23



PRIOR APPLICATION NUMBER: US 60/156,358  
PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: US 60/146,002  
PRIOR FILING DATE: 1999-08-09  
NUMBER OF SEQ ID NOS: 325720  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 8782  
LENGTH: 737  
TYPE: DNA  
ORGANISM: Human  
US-10-027-632-8782

Query Match 13.2%; Score 39.6; DB 15; Length 737;  
Best Local Similarity 66.3%; Pred. No. 0.19;  
Matches 57; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 98 TAGCTGCATTGCTTGGCATTCCTAGCAGCCTGGCATAGTTCCATTACAGTACGC 157  
DB 385 TATCTATTGCTCTCATATGATGATCCTGAGCATCGAAGACAGCTGTGCTTGCATTAGGC 326  
QY 158 ATGAAGTATCTCTCAGTGAATTAAT 183  
DB 325 ACTCAGTATTCTACTGATGAATTAAT 300

RESULT 10  
US-09-764-872-490  
Sequence 490, Application US/09764872  
Publication No. US20030050231A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PA125  
CURRENT APPLICATION NUMBER: US/09/764,872  
CURRENT FILING DATE: 2001-01-17  
Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 957  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 490  
LENGTH: 11216  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-764-872-490

Query Match 13.2%; Score 39.6; DB 10; Length 11216;  
Best Local Similarity 66.3%; Pred. No. 0.75;  
Matches 57; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 98 TAGCTGCATTGCTTGGCATTCCTAGCAGCCTGGCATAGTTCCATTACAGTACGC 157  
DB 10173 TATCTATTGCTCTCATATGATGATCCTGAGCATCGAAGACAGCTGTGCTTGCATTAGGC 10232  
QY 158 ATGAAGTATCTCTCAGTGAATTAAT 183  
DB 10233 ACTCAGTATTCTACTGATGAATTAAT 10258

RESULT 11  
US-10-312-841-2/c  
Sequence 2, Application US/10312841  
Publication No. US20030186277A1  
GENERAL INFORMATION:  
APPLICANT: Epigenomics AG  
TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC  
FILE REFERENCE: E01/1208/NO  
CURRENT APPLICATION NUMBER: US/10/312,841  
CURRENT FILING DATE: 2002-12-30  
NUMBER OF SEQ ID NOS: 2  
SEQ ID NO 2  
LENGTH: 3673778  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:

OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (379615)  
US-10-312-841-2

Query Match 12.5%; Score 37.6; DB 14; Length 3673778;  
Best Local Similarity 61.0%; Pred. No. 55;  
Matches 61; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 177 AATTAATGATTCATATGGGCTTACAGTACGAGAGAGAGATTAATCTTAATAGCCAA 236  
DB 50379 AAAAAACAATTCATATCTTTAATTCGAAAAATATATAATTAATATACCA 50320  
QY 237 GTCTATGAGAGACAAATTAAGTCTTTAATTTGAGA 276  
DB 50319 TTAATAATACCATTAATAACAACTAATATACAAA 50280

RESULT 12  
US-10-029-386-5897  
Sequence 5897, Application US/10029386  
Publication No. US20030194704A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: ABOVICA-X-2  
CURRENT APPLICATION NUMBER: US/10/029,386  
CURRENT FILING DATE: 2001-12-20  
NUMBER OF SEQ ID NOS: 34286  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 5897  
LENGTH: 600  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AF216808.2  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.98  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.98  
OTHER INFORMATION: NT HIT: AL163267.2, EVALUE 1.50e-01  
OTHER INFORMATION: EST\_HUMAN HIT: BE243894.1, EVALUE 8.40e-01  
US-10-029-386-5897

Query Match 12.5%; Score 37.4; DB 14; Length 600;  
Best Local Similarity 54.8%; Pred. No. 0.77;  
Matches 74; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 88 GCAAGACCTTACCTGATTCCTAGCATTCCTAGCAGCCTGGCATAGTTCCATT 147  
DB 279 GCAAGAGTAATATGCTTAGCATCTTGAATCCCTAGTACCTAGCAATGCTTAAT 338  
QY 148 AACAGTGGCATGATGATCTACTCAGTGAATTAATGAATGATATGGCTTACAGTAG 207  
DB 339 CTTAGTAGGATGATGATGCTTATTAATAAATAATCATGATGATTAATAACTGAT 398  
QY 208 AGAGAGAAATTAAT 222  
DB 399 AATTTGAGATTAAT 413

RESULT 13  
US-10-027-632-270562/c  
Sequence 270562, Application US/10027632  
Publication No. US20030204075A9  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
TITLE OF INVENTION: Polymorphisms in the Human Genome  
FILE REFERENCE: 108827.129



```

RESULT 14
US-10-257-166-47/c
; Sequence 47, Application US/10257166
; Publication No. US20040023230A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Method and Nucleic Acids for Analysing the Methylation of
; TITLE OF INVENTION: Genes Implicated in Pharmacogenomics
; FILE REFERENCE: 5013.1011
; CURRENT APPLICATION NUMBER: US/10/257,166
; CURRENT FILING DATE: 2002-10-07
; PRIOR APPLICATION NUMBER: PCT/EP01/07470
; DE 10032529.7
; DE 10043826.1
; PRIOR FILING DATE: 2001-06-29
; 2000-06-30
; 2000-09-01
; NUMBER OF SEQ ID NOS: 178
; SEQ ID NO 47
; LENGTH: 8943
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-257-166-47

```

Query Match	Best Local Similarity	Score	DB	Length
Matches	53; Conservative	12.1%; 65.4%;	DB 15;	637;
			Pred. No. 1.8;	
			Mismatches 28;	Indels 0;
			Gaps 0	
QY	68	AGGCTTGAGACGAGAAAGAGCAGACCCCTGCTGCTATGCTTATGATGCATCCCTAG	127	
DB	469	AGATGTAAAGCTCCAGAGAGGCGAAACAGTCTCTTACCTTGTATAGTCGATTCCTTAG	410	
QY	128	CACCTGGCATAGTTTCCATTTA	148	
DB	409	CACCTGGCATGATGTTCTTGT	389	







ORGANISM: Nucleotide sequence of HC-contig  
US-09-078-294-3

Query Match 30.5%; Score 61.4; DB 3; Length 80595;  
Best Local Similarity 86.1%; Pred. No. 4.7e-09;  
Matches 68; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 120 CTCCTCTCTCTCCACACACACACACACACACACACACACACACACAA 179  
DB 52316 CTCCTCTCTCTGACACACACACACACACACACACACACACACACAC 52375  
QY 180 ACACACACCCCGCCACCA 198  
DB 52376 ACACACACACACACACACACAC 52394

## RESULT 3

US-08-222-177A-14  
Sequence 14, Application US/08222177A

Patent No. 5582979

GENERAL INFORMATION:

APPLICANT: Weber, James L.

TITLE OF INVENTION: LENGTH POLYMORPHISMS IN

TITLE OF INVENTION: (dc-da)n.(dg-dt)n SEQUENCES AND METHODS OF USING SAME

NUMBER OF SEQUENCES: 460

CORRESPONDENCE ADDRESS:

ADDRESSEE: Dewitt Ross & Stevens, S.C.

STREET: 8000 Excelsior Drive, Suite 401

CITY: Madison

STATE: Wisconsin

COUNTRY: USA

ZIP: 53717-1914

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/222.177A

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/341,562

FILING DATE: 21-APR-1989

ATTORNEY/AGENT INFORMATION:

NAME: Sara, Charles S.

REGISTRATION NUMBER: 30,492

REFERENCE/DOCKET NUMBER: 09865, 601

TELECOMMUNICATION INFORMATION:

TELEPHONE: (608) 831-2100

TELEFAX: (608) 831-2106

TELEX:

INFORMATION FOR SEQ. ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 223 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

INDIVIDUAL ISOLATE: Caucasian

TISSUE TYPE: Blood

IMMEDIATE SOURCE:

CLONE: M126

POSITION IN GENOME:

CHROMOSOME/SEGMENT: 18

FEATURE:

NAME/KEY: repeat region

LOCATION: 129..185

OTHER INFORMATION: /rpt\_type=" tandem"

OTHER INFORMATION: /rpt\_family=" (dc-da)n.(dg-dt)n"

OTHER INFORMATION: /citation= (12)

FEATURE:

NAME/KEY: misc feature

LOCATION: 90..109

IDENTIFICATION METHOD: experimental

OTHER INFORMATION: /evidence= EXPERIMENTAL

OTHER INFORMATION: /standard\_name= "PCR primer"

OTHER INFORMATION: /citation= (11)

FEATURE:

NAME/KEY: misc feature

LOCATION: complement (185..204)

IDENTIFICATION METHOD: experimental

OTHER INFORMATION: /evidence= EXPERIMENTAL

OTHER INFORMATION: /standard\_name= "PCR primer"

OTHER INFORMATION: /citation= (11)

FEATURE:

NAME/KEY: misc feature

LOCATION: 1..223

IDENTIFICATION METHOD: experimental

OTHER INFORMATION: /evidence= EXPERIMENTAL

OTHER INFORMATION: /standard\_name= "Only one strand sequenced"

OTHER INFORMATION: /citation= (11)

FEATURE:

NAME/KEY: misc feature

LOCATION: 1..223

IDENTIFICATION METHOD: experimental

OTHER INFORMATION: /evidence= EXPERIMENTAL

OTHER INFORMATION: /standard\_name= "Only one strand sequenced"

OTHER INFORMATION: /citation= (11)

FEATURE:

NAME/KEY: misc feature

LOCATION: 1..223

IDENTIFICATION METHOD: experimental

OTHER INFORMATION: /evidence= EXPERIMENTAL

OTHER INFORMATION: /standard\_name= "Only one strand sequenced"

OTHER INFORMATION: /citation= (11)

FEATURE:

NAME/KEY: misc feature

LOCATION: 1..223

IDENTIFICATION METHOD: experimental

OTHER INFORMATION: /evidence= EXPERIMENTAL

OTHER INFORMATION: /standard\_name= "Only one strand sequenced"

OTHER INFORMATION: /citation= (11)

FEATURE:

NAME/KEY: misc feature

LOCATION: 1..223

IDENTIFICATION METHOD: experimental

OTHER INFORMATION: /evidence= EXPERIMENTAL

OTHER INFORMATION: /standard\_name= "Only one strand sequenced"

OTHER INFORMATION: /citation= (11)

FEATURE:

NAME/KEY: misc feature

LOCATION: 1..223

IDENTIFICATION METHOD: experimental

OTHER INFORMATION: /evidence= EXPERIMENTAL

OTHER INFORMATION: /standard\_name= "Only one strand sequenced"

OTHER INFORMATION: /citation= (11)

FEATURE:

NAME/KEY: misc feature

LOCATION: 1..223

IDENTIFICATION METHOD: experimental

OTHER INFORMATION: /evidence= EXPERIMENTAL

OTHER INFORMATION: /standard\_name= "Only one strand sequenced"

OTHER INFORMATION: /citation= (11)

FEATURE:

NAME/KEY: misc feature

LOCATION: 1..223

IDENTIFICATION METHOD: experimental

OTHER INFORMATION: /evidence= EXPERIMENTAL

OTHER INFORMATION: /standard\_name= "Only one strand sequenced"

OTHER INFORMATION: /citation= (11)

## RESULT 4

US-09-146-053-4/C  
Sequence 4, Application US/09146053A

Patent No. 6399349

GENERAL INFORMATION:

APPLICANT: Ryan, James W.

APPLICANT: Sprinkle, Terry Joe Curtis

APPLICANT: Venema, Richard C.

TITLE OF INVENTION: Human Aminopeptidase P Gene

FILE REFERENCE: MCG103

CURRENT APPLICATION NUMBER: US/09/146,053A

CURRENT FILING DATE: 1998-09-02

EARLIER APPLICATION NUMBER: 60/057,854

EARLIER FILING DATE: 1997-09-02

NUMBER OF SEQ ID NOS: 7

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 4

LENGTH: 50000

TYPE: DNA

US-08-222-177A-14

Query Match 29.8%; Score 59.8; DB 1; Length 223;

Best Local Similarity 71.2%; Pred. No. 2.2e-09;

Matches 79; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 89 TCACGACACTTTTCTTCTTATGATAGGCTCTCTCTCCACACACACACCA 148  
DB 88 TTCAGAAATTCCTCTGCTATTTTCAATTATTTCTTGAACACACACACACA 147  
QY 149 CACCA 199  
DB 148 CACCA 198















```

SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/819,458A
FILING DATE: 17-MAR-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: lambdis, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 5215,000-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO.: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4157 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-819-458A-1

```

Query Match	28.7%	Score 57.6; DB 2;	Length 4157;
Best Local Similarity	87.5%	Pred. No. 2.6e-08;	
Matches 63; Conservative	0;	Mismatches 9;	Incls 0; Gaps 0;

QY	TCTCTCCACACACACACACACACACACACACACACACACAA	186
Dbb	TCATTACACACACACACACACACACACACACACACACACA	348-9

QY	187	CCCCGCCACCA	198
Db	3485	CACACACACACA	3496

RESULT 15  
US-09-018-864A-1

```

; Sequence 1, Application US/09018866A
; Patent No. 5958747
; GENERAL INFORMATION:

```

APPLICANT: Elrod, Susan L.  
APPLICANT: Cherry, Joel R.  
TITLE OF INVENTION: Aspergillus Oryzae 5-Aminolevulinic  
TITLE OF INVENTION: Acid Synthases And Nucleic Acids Encoding Same  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESS: NO. 59587470 NO. 59587470 OF NO. 59587470h America, Inc  
STREET: 405 Lexington Avenue - 64th Fl.

```

; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
;

```

```

?      COMPUTER READABLE FORM:
?
?      MEDIUM TYPE: Diskette
?
?      COMPUTER: IBM Compatible
?
?      OPERATING SYSTEM: DOS
?
?      SOFTWARE: FastSeq for Windows Version 2.0
?

```

;;  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/018,864A  
FILING DATE:  
;

```

; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/871,266
;

```

;  
;  
;  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lambiris, Elias J

REGISTRATION NUMBER: 33,728  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-878-9652  
TELEFAX: 212-878-9655  
TELEX:

```

; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 4157 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-018-864A-1

```

Query Match	28.7%	Score 57.6;	DB 2;	Length 4157;
Best Local Similarity	87.5%	Pred. No. 2.6e-08;		
Matches 63;	Conservative 0;	Mismatches 9;	Indels 0;	Gaps 0;

Dy  
Db

127 TCTGTCCACACA CACACACACA CACACACACA CACACACACA CAACAAACACACA 186  
3425 TCATATACACACA CACACACACA CACACACACA CACACACACA CACACACACA 3486

QY	187	CCCCGCCAACCA	198
Db	3485	CACACACACACA	3496

Search completed: March 12, 2004, 07:24:00  
Job time : 31.8242 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 11, 2004, 17:19:49 ; Search time 352.627 Seconds  
(without alignments)

2098.537 Million cell updates/sec

Title: US-09-966-880A-9\_COPY\_2707\_2907  
Perfect score: 201  
Sequence: 1 gtagagcttattatgtrttt.....acacaccgcgcacacaag 201

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2432557 seqs, 1840798884 residues

Total number of hits satisfying chosen parameters: 4865114

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEM\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEM\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEM\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEM\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEM\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10\_NEM\_PUB.seq:\*
- 17: /cgn2\_6/ptodata/2/pubpna/US60\_NEM\_PUB.seq:\*
- 18: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	201	100.0	5514	9	US-09-966-880A-9
2	201	100.0	11204	9	US-09-966-880A-35
3	70	34.8	3576	11	US-09-245-277-37
4	70	34.8	3578	11	US-09-244-805-37
5	69.4	34.5	6927	15	US-10-429-681-4
6	67.2	33.4	6427	15	US-10-027-632-261170
7	66.4	33.0	165961	15	US-10-085-117-241
8	65.2	32.4	15306	14	US-10-238-676-162
9	65.2	32.4	15306	14	US-10-240-453-184
10	64.8	32.2	406	15	US-09-233-785-256
11	64	31.8	525	15	US-10-027-632-45364
12	64	31.8	605	15	US-10-027-632-275968
13	64	31.8	630	15	US-10-027-632-301462
14	63.8	31.7	362	15	US-10-027-632-147955
15	63.8	31.7	719	15	US-10-027-632-250183

16	63.6	31.6	462	15	US-10-027-632-275967	Sequence 275967, A
17	63.6	31.6	468	10	US-09-918-995-14728	Sequence 14728, A
18	63.6	31.6	5179	14	US-10-311-455-1969	Sequence 1969, Ap
19	63.6	31.6	5179	14	US-10-240-453-277	Sequence 277, App
20	63.2	31.4	82938	10	US-09-818-657-3	Sequence 3, Appli
21	63	31.3	80246	10	US-09-728-552-4	Sequence 4, Appli
22	63	31.3	263744	14	US-10-229-834A-6	Sequence 6, Appli
23	62.8	31.2	878	15	US-10-027-632-119922	Sequence 119922, A
24	62.6	31.1	362	15	US-10-027-632-20442	Sequence 20442, A
25	62.6	31.1	215980	10	US-09-972-546-16	Sequence 16, Appli
26	62.4	31.0	26042	14	US-10-004-113-31	Sequence 31, Appli
27	62.2	30.9	887	15	US-10-027-632-164621	Sequence 164621, A
28	62.2	30.9	26042	14	US-10-027-632-164622	Sequence 164622, A
29	61.8	30.7	26042	14	US-10-004-113-31	Sequence 31, Appli
30	61.4	30.5	80595	10	US-09-728-552-3	Sequence 3, Appli
31	60.8	30.2	481	15	US-10-027-632-298608	Sequence 298608, A
32	60.8	30.2	96602	15	US-10-085-117-61	Sequence 61, Appli
33	60.6	30.1	478	15	US-10-027-632-83629	Sequence 83629, A
34	60.6	30.1	478	15	US-10-027-632-83630	Sequence 83630, A
35	60.6	30.1	659158	9	US-09-771-208-20	Sequence 20, Appli
36	60.4	30.0	575	15	US-10-027-632-279794	Sequence 279794, A
37	60.2	30.0	452	9	US-09-864-761-651	Sequence 651, App
38	60	29.9	798	15	US-10-027-632-128100	Sequence 128100, A
39	60	29.9	798	15	US-10-027-632-128101	Sequence 128101, A
40	60	29.9	1067	15	US-10-027-632-264022	Sequence 264022, A
41	60	29.9	1067	15	US-10-027-632-264023	Sequence 264023, A
42	60	29.9	8995	9	US-09-877-935-1	Sequence 1, Appli
43	59.8	29.8	15698	14	US-10-311-455-2114	Sequence 2114, Ap
44	59.8	29.8	169139	14	US-10-067-514-1	Sequence 1, Appli
45	59.8	29.8	169139	15	US-10-419-723-1	Sequence 1, Appli

## ALIGNMENTS

RESULT 1  
US-09-966-880A-9  
Sequence 9, Application US/09966880A  
Patent No. US20020164743A1  
GENERAL INFORMATION:  
APPLICANT: Honjo, Tasuku  
TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
FILE REFERENCE: 06501-088001 US/09/966, 880A  
CURRENT APPLICATION NUMBER: PCT/JP00/01918  
CURRENT FILING DATE: 2001-09-28  
PRIOR APPLICATION NUMBER: US/09-966, 880A  
PRIOR FILING DATE: 2000-03-28  
PRIOR APPLICATION NUMBER: JP 11-371382  
PRIOR FILING DATE: 1999-12-27  
PRIOR APPLICATION NUMBER: JP 11-178999  
PRIOR FILING DATE: 1999-06-24  
PRIOR APPLICATION NUMBER: JP 11-87192  
PRIOR FILING DATE: 1999-03-29  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 9  
LENGTH: 5514  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: Intron  
LOCATION: (1) ... (1031)  
FEATURE:  
NAME/KEY: Exon  
LOCATION: (1032) ... (1118)  
FEATURE:  
NAME/KEY: Intron  
LOCATION: (1119) ... (5514)  
US-09-966-880A-9

Query Match 100.0%; Score 201; DB 9; Length 5514;  
Best local similarity 100.0%; Pred. No. 1.4e-51;



Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGAGGCTTATGTAATTTTCATGAGTAGAGAGCCCAAAATCCACCAGTCTTTA 60  
DB 2707 GTGAGGCTTATGTAATTTTCATGAGTAGAGAGCCCAAAATCCACCAGTCTTTA 2766

QY 61 TCTATGCCACATCCTCTTATCTATCTTCCAGAGCACTTTTCTCTCTTATGATAAGC 120  
DB 2767 TCTATGCCACATCCTCTTATCTATCTTCCAGAGCACTTTTCTCTCTTATGATAAGC 2826

QY 121 TCTCTCTCTCCAA 180  
DB 2827 TCTCTCTCTCTCCACACACACACACACACACACACACACACACACACACAAA 2886

QY 181 CACACACCCCGCCCAACCAAG 201  
DB 2887 CACACACCCCGCCCAACCAAG 2907

RESULT 2  
US-09-966-880A-35  
Sequence 35, Application US/09966880A  
Patent No. US2002016473A1  
GENERAL INFORMATION:  
APPLICANT: Honjo, Tasuku  
APPLICANT: Muramatsu, Masaamichi  
TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
FILE REFERENCE: 06501-088001  
CURRENT APPLICATION NUMBER: US/09/966, 880A  
PRIOR FILING DATE: 2001-09-28  
PRIOR APPLICATION NUMBER: PCT/JP00/01918  
PRIOR FILING DATE: 2000-03-28  
PRIOR APPLICATION NUMBER: JP 11-371382  
PRIOR FILING DATE: 1999-12-27  
PRIOR APPLICATION NUMBER: JP 11-178999  
PRIOR FILING DATE: 1999-06-24  
PRIOR APPLICATION NUMBER: JP 11-87192  
PRIOR FILING DATE: 1999-03-29  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 35  
LENGTH: 11204  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-966-880A-35

Query Match 100.0%; Score 201; DB 9; Length 11204;  
Best Local Similarity 100.0%; Pred. No. 1.8e-51;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGAGGCTTATGTAATTTTCATGAGTAGAGAGCCCAAAATCCACCAGTCTTTA 60  
DB 2117 GTGAGGCTTATGTAATTTTCATGAGTAGAGAGCCCAAAATCCACCAGTCTTTA 2176

QY 61 TCTAGGCCACATCCTCTTATCTATCTTCCAGAGCACTTTTCTCTCTTATGATAAGC 120  
DB 2177 TCTAGGCCACATCCTCTTATCTATCTTCCAGAGCACTTTTCTCTCTTATGATAAGC 2236

QY 121 TCTCTCTCTCCAA 180  
DB 2237 TCTCTCTCTCTCCACACACACACACACACACACACACACACACACACACAAA 2286

QY 181 CACACACCCCGCCCAACCAAG 201  
DB 2297 CACACACCCCGCCCAACCAAG 2317

RESULT 3  
US-09-245-277-37/c  
Sequence 37, Application US/09245277  
Publication No. US20030211984A1  
GENERAL INFORMATION:  
APPLICANT: Morley, Paul F.

APPLICANT: Lashan, Anthony  
TITLE OF INVENTION: IMMEDIATE EARLY GENES AND METHODS OF USE  
FILE REFERENCE: JH1530-3  
CURRENT APPLICATION NUMBER: US/09/245, 277  
CURRENT FILING DATE: 1999-02-05  
PRIOR APPLICATION NUMBER: 60/074, 518  
PRIOR FILING DATE: 1998-02-12  
PRIOR APPLICATION NUMBER: 60/074, 135  
PRIOR FILING DATE: 1998-02-06  
NUMBER OF SEQ ID NOS: 63  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 37  
LENGTH: 3976  
TYPE: DNA  
ORGANISM: Eukaryote  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (51)...(1790)  
US-09-245-277-37

Query Match 34.8%; Score 70; DB 11; Length 3976;  
Best Local Similarity 72.2%; Pred. No. 2.7e-11;  
Matches 91; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 69 ACATCCTCTTATCTATCTTCCAGAGCACTTTTCTCTCTTATGATAAGGCTCTCTC 128  
DB 3100 ACATCCTC 3041

QY 129 TCTCCACC 188  
DB 3040 TCTACA 2981

QY 189 CCGCCA 194  
DB 2980 CACACA 2975

RESULT 4  
US-09-244-805-37/c  
Sequence 37, Application US/09244805  
Publication No. US20030203840A1  
GENERAL INFORMATION:  
APPLICANT: Morley, Paul F.  
APPLICANT: Lashan, Anthony  
APPLICANT: Goetz, Bernard  
APPLICANT: Heimsch, Holger  
APPLICANT: Kuner, Rohini  
APPLICANT: Schenk, Sigrid  
APPLICANT: Nikolich, Karoly  
APPLICANT: Zhukovskii, Eugene  
TITLE OF INVENTION: IMMEDIATE EARLY GENES AND METHODS OF USE  
FILE REFERENCE: 10496/004001  
CURRENT APPLICATION NUMBER: US/09/244, 805  
CURRENT FILING DATE: 1999-02-05  
EARLIER APPLICATION NUMBER: 60/074, 518  
EARLIER FILING DATE: 1998-02-12  
EARLIER APPLICATION NUMBER: 60/074, 135  
EARLIER FILING DATE: 1998-02-06  
NUMBER OF SEQ ID NOS: 62  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 37  
LENGTH: 3976  
TYPE: DNA  
ORGANISM: Eukaryote  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (51)...(1790)  
US-09-244-805-37

Query Match 34.8%; Score 70; DB 11; Length 3978;  
Best Local Similarity 72.2%; Pred. No. 2.7e-11;



Matches 91; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 69 ACATCTCTTACTTACTTCCAGACACTTTTCTTCTTATGATAGGCTCTCTC 128  
DB 3102 ACAACTC 3043

QY 129 TCTCCAC 188  
DB 3042 TCTAC 2983

QY 189 CCGCCA 194  
DB 2982 CACACA 2977

RESULT 5  
US-10-429-681-4/c

Sequence 4, Application US/10429681  
Publication No. US20040003425A1  
GENERAL INFORMATION:  
APPLICANT: Japan as Represented by Director General of Okazaki National Research  
APPLICANT: Institutes  
TITLE OF INVENTION: Nav2 channel gene-deficient non-human animals  
FILE REFERENCE: U2001P059  
CURRENT APPLICATION NUMBER: US/10/429,681  
CURRENT FILING DATE: 2003-05-06  
PRIOR APPLICATION NUMBER: JP 2000/237320  
PRIOR FILING DATE: 2000-08-04  
PRIOR APPLICATION NUMBER: JP 2000/241637  
PRIOR FILING DATE: 2000-08-09  
PRIOR APPLICATION NUMBER: JP 2001/222263  
PRIOR FILING DATE: 2001-07-23  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 4  
LENGTH: 6927  
TYPE: DNA  
ORGANISM: Mus musculus  
US-10-429-681-4

Query Match 34.5%; Score 69.4; DB 15; Length 6927;  
Best Local Similarity 92.4%; Pred. No. 5.1e-11;  
Matches 73; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 120 CT 179  
DB 3038 CT 2979

QY 180 ACACACACCCCGCAACCA 198  
DB 2978 ACACACACCACTACCA 2960

RESULT 6  
US-10-027-632-261170  
Sequence 261170, Application US/10027632  
Publication No. US20030204075A9  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
FILE REFERENCE: 108827.129  
CURRENT APPLICATION NUMBER: US/10/027,632  
CURRENT FILING DATE: 2002-04-30  
PRIOR APPLICATION NUMBER: US 60/218,006  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/198,676  
PRIOR FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: US 60/193,483  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: US 60/185,218  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/167,363

PRIOR FILING DATE: 1999-11-23  
PRIOR APPLICATION NUMBER: US 60/156,358  
PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: US 60/146,002  
PRIOR FILING DATE: 1999-08-09  
NUMBER OF SEQ ID NOS: 325720  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 261170  
LENGTH: 647  
TYPE: DNA  
ORGANISM: Human  
US-10-027-632-261170

Query Match 33.4%; Score 67.2; DB 15; Length 647;  
Best Local Similarity 85.2%; Pred. No. 1e-10;  
Matches 75; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 100 TTTTCTCTTATGATAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 159  
DB 73 TTGTTTCTCTGATGATGCGCATGACTCCCTGCACACACACACACACACAC 132

QY 160 AC 187  
DB 133 AC 160

RESULT 7  
US-10-085-117-241

Sequence 241, Application US/10085117  
Publication No. US2003023334A1  
GENERAL INFORMATION:  
APPLICANT: Morris, David W.  
TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER  
FILE REFERENCE: 529452000121  
CURRENT APPLICATION NUMBER: US/10/085,117  
CURRENT FILING DATE: 2002-02-27  
PRIOR APPLICATION NUMBER: US 09/798,586  
PRIOR FILING DATE: 2001-03-02  
NUMBER OF SEQ ID NOS: 361  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 241  
LENGTH: 165961  
TYPE: DNA  
ORGANISM: Mus musculus  
FEATURE:  
NAME/KEY: variation  
LOCATION: (1)...(165961)  
OTHER INFORMATION: n = any nucleotide  
US-10-085-117-241

Query Match 33.0%; Score 66.4; DB 15; Length 165961;  
Best Local Similarity 73.3%; Pred. No. 1.4e-09;  
Matches 85; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 78 TTATCTTACTTCCAGACACTTTTCTTCTTATGATAGGCTCTCTCTCTCT 137  
DB 2691 TGACTTATCAGTTAGGCGCACTTGTTCTTTCACAGCTGGGTTCCAGTCC 2750

QY 138 AC 193  
DB 2751 ACATGAC 2806

## RESULT 8

US-10-239-676-162/c  
Sequence 162, Application US/10239676  
Publication No. US2003082609A1  
GENERAL INFORMATION:  
APPLICANT: OLEK, Alexander  
APPLICANT: PREPENEROCK, Christian  
TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation



FILE REFERENCE: 5013.1003  
CURRENT APPLICATION NUMBER: US/10/239,676  
CURRENT FILING DATE: 2002-09-24  
PRIOR APPLICATION NUMBER: PCT/EP01/03968  
DE 10019058.8  
DE 10019173.8  
DE 10032529.7  
DE 10043826.1  
PRIOR FILING DATE: 2001-04-06  
2000-04-06  
2000-04-07  
2000-06-30  
2000-09-01  
NUMBER OF SEQ ID NOS: 228  
SEQ ID NO 162  
LENGTH: 15306  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-239-676-162

Query Match 32.4%; Score 65.2; DB 14; Length 15306;  
Best Local Similarity 72.0%; Pred. No. 1.3e-09;  
Matches 85; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 82 CTATACCTCCAGACACTTTTCTCTCTATAGTAAAGCTCTCTCTCTCCACACAC 141  
DB 10351 CTAAATTTTCACCACTATTTCCTCCCTTAAACAACTATCACTCAACACAC 10292  
QY 142 ACAAC 199  
DB 10291 ACAAC 10234

## RESULT 9

US-10-240-453-184/c  
Sequence 184, Application US/10240453  
Publication No. US20030148326A1  
GENERAL INFORMATION:  
APPLICANT: OLEK, Alexander  
APPLICANT: PIEPENBROCK, Christian  
APPLICANT: BERLIN, Kurt  
TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA  
TITLE OF INVENTION: by Means of Assessing the Methylation Status of Genes Associated  
TITLE OF INVENTION: With DNA Transcription  
FILE REFERENCE: 5013.1009  
CURRENT APPLICATION NUMBER: US/10/240,453  
CURRENT FILING DATE: 2002-10-02  
PRIOR APPLICATION NUMBER: PCT/EP01/03973  
PRIOR FILING DATE: 2001-04-06  
PRIOR APPLICATION NUMBER: DE 10019058.8  
PRIOR FILING DATE: 2000-04-06  
PRIOR APPLICATION NUMBER: DE 10019173.8  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: DE 10032529.7  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: DE 10043826.1  
PRIOR FILING DATE: 2000-09-01  
NUMBER OF SEQ ID NOS: 350  
SEQ ID NO 184  
LENGTH: 15306  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-240-453-184

Query Match 32.4%; Score 65.2; DB 14; Length 15306;  
Best Local Similarity 72.0%; Pred. No. 1.3e-09;  
Matches 85; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 82 CTATACCTCCAGACACTTTTCTCTCTATAGTAAAGCTCTCTCTCTCCACACAC 141  
DB 10351 CTAAATTTTCACCACTATTTCCTCCCTTAAACAACTATCACTCAACACAC 10292  
QY 142 ACAAC 199  
DB 10291 ACAAC 10234

## RESULT 10

US-09-232-785-256/c  
Sequence 256, Application US/09232785  
Publication No. US20030049612A1  
GENERAL INFORMATION:  
APPLICANT: International Paper Co.  
APPLICANT: Eght, Craig S  
APPLICANT: Nelson, C. Dana  
TITLE OF INVENTION: MICROSATellite DNA MARKERS AND USES  
TITLE OF INVENTION: THEREOF  
FILE REFERENCE: 4481/1E18US1  
CURRENT APPLICATION NUMBER: US/09/232,785  
CURRENT FILING DATE: 1999-01-19  
PRIOR APPLICATION NUMBER: 09/232,884  
PRIOR FILING DATE: 1999-01-15  
NUMBER OF SEQ ID NOS: 397  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 256  
LENGTH: 406  
TYPE: DNA  
ORGANISM: Pinus taeda L.  
FEATURE:  
NAME/KEY: unsure  
LOCATION: n at 43 and 406.  
OTHER INFORMATION: n is a or g or c or t/u, unknown or other.  
US-09-232-785-256

Query Match 32.2%; Score 64.8; DB 10; Length 406;  
Best Local Similarity 90.8%; Pred. No. 4.8e-10;  
Matches 69; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 112 TGATAGGCTCTCTCTCTCTCCACACACACACACACACACACACACACACAC 171  
DB 357 TGAGATGATCTCTCTCTCTATACACACACACACACACACACACACACACAC 298  
QY 172 ACAACAAACACACAC 187  
DB 297 ACAACACACACACAC 282

## RESULT 11

US-10-027-632-45364/c  
Sequence 45364, Application US/10027632  
Publication No. US20030204075A9  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
TITLE OF INVENTION: Polymorphisms in the Human Genome  
FILE REFERENCE: 108827.129  
CURRENT APPLICATION NUMBER: US/10/027,632  
CURRENT FILING DATE: 2002-04-30  
PRIOR APPLICATION NUMBER: US 60/218,006  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/198,676  
PRIOR FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: US 60/193,483  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: US 60/185,218  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/167,363  
PRIOR FILING DATE: 1999-11-23  
PRIOR APPLICATION NUMBER: US 60/156,358  
PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: US 60/146,002



```

?
? PRIOR FILING DATE: 1999-08-09
?
? NUMBER OF SEQ ID NOS: 325720
?
? SOFTWARE: FASTSEQ for Windows Version 4.0
?
? SEQ ID NO: 45364
?
? LENGTH: 525
?
? TYPE: DNA
?
? ORGANISM: Human
?
US-10-027-632-45364

```

Query Match	31.8%;	Score 64;	DB 15;	Length 525;
Best Local Similarity	60.6%;	Pred. No. 9.2e-10;		
Matches 103; Conservative	1;	Mismatches 66;	Indels 0;	Gaps 0;

[illegible]

RESULT 12  
US-10-027-632-275968

```

1. APPLICANT: Wang, David G.
2. TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
3. TITLE OF INVENTION: Polymorphisms in the Human Genome
4. FILE REFERENCE: 108827, 129
5. CURRENT APPLICATION NUMBER: US/10/027,632
6. CURRENT FILING DATE: 2002-04-30
7. PRIOR APPLICATION NUMBER: US 60/218,006
8. PRIOR FILING DATE: 2000-07-12
9. PRIOR APPLICATION NUMBER: US 60/198,676
10. PRIOR FILING DATE: 2000-04-20
11. PRIOR APPLICATION NUMBER: US 60/193,483
12. PRIOR FILING DATE: 2000-03-29
13. PRIOR APPLICATION NUMBER: US 60/185,218
14. PRIOR FILING DATE: 2000-02-24
15. PRIOR APPLICATION NUMBER: US 60/167,363
16. PRIOR FILING DATE: 1999-11-23
17. PRIOR APPLICATION NUMBER: US 60/156,358
18. PRIOR FILING DATE: 1999-09-28
19. PRIOR APPLICATION NUMBER: US 60/146,002
20. PRIOR FILING DATE: 1999-08-09
21. NUMBER OF SEQ ID NOS: 325720
22. SOFTWARE: FastSeq for Windows Version 4.0
23. SEQ ID NO 275968
24. LENGTH: 605
25. TYPE: DNA
26. ORGANISM: Human
27. US-10-027-632-275968

```

```

Query Match      31.8%; Score 64; DB 15; Length 605;
Best Local Similarity 93.1%; Pred. No. 9.7e-10;
Matches 67; Conservative 0; Mismatches 5; Indels 0; Gaps 0

Oy      118 GGGCTTCTCTCTCTGCACACACACACACACACACACACACACAC 177
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      332 GTCTCTCTCTCTCTTCATTAACACACACACACACACACACACACACAC 451

Oy      178 AAACACACACCC 189
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      452 AACACACACAC 463

```

## RESULT 13

```
US-10-027-632-301462/c
; Sequence 301462, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match	31.8%;	Score 64;	DB 15;	Length 630;
Best Local Similarity	60.6%;	Pred. No. 9.8e-10;		
Matches 103; Conservative	1;	Mismatches 66;	Indels 0;	Gaps 0

[illegible]

RESULT 14  
US-10-027-632-147955/c  
Sequence 147955, Application US/10027632  
Publication No. US20030204075A9  
GENERAL INFORMATION:  
APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
Polymorphisms in the Human Genome  
FILE REFERENCE: 108827, 129  
CURRENT APPLICATION NUMBER: US/10/027,632  
CURRENT FILING DATE: 2002-04-30  
PRIOR APPLICATION NUMBER: US 60/218,006  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: US 60/198,676  
PRIOR FILING DATE: 2000-04-20  
PRIOR APPLICATION NUMBER: US 60/193,483  
PRIOR FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: US 60/185,218  
PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/167,363  
PRIOR FILING DATE: 1999-11-23  
PRIOR APPLICATION NUMBER: US 60/156,356  
PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: US 60/146,002

```

1  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
2  TITLE OF INVENTION: Polymorphisms in the Human Genome
3  FILE REFERENCE: 108827.129
4  CURRENT APPLICATION NUMBER: US/1,0/027,632
5  CURRENT FILING DATE: 2002-04-30
6  PRIOR APPLICATION NUMBER: US 60/218,006
7  PRIOR FILING DATE: 2000-07-12
8  PRIOR APPLICATION NUMBER: US 60/198,676
9  PRIOR FILING DATE: 2000-04-20
10 PRIOR APPLICATION NUMBER: US 60/193,483
11 PRIOR FILING DATE: 2000-03-29
12 PRIOR APPLICATION NUMBER: US 60/185,218
13 PRIOR FILING DATE: 2000-02-24
14 PRIOR APPLICATION NUMBER: US 60/167,363
15 PRIOR FILING DATE: 1999-11-23
16 PRIOR APPLICATION NUMBER: US 60/156,358
17 PRIOR FILING DATE: 1999-09-28
18 PRIOR APPLICATION NUMBER: US 60/146,002

```



Job time : 355.627 secs

/ PRIOR FILING DATE: 1999-08-09  
/ NUMBER OF SEQ ID NOS: 325720  
/ SOFTWARE: FastSeq for Windows Version 4.0  
/ SEQ ID NO 147955  
/ LENGTH: 362  
/ TYPE: DNA  
/ ORGANISM: Human  
US-10-027-632-147955

Query Match 31.7%; Score 63.8; DB 15; Length 362;  
Best Local Similarity 88.3%; Pred. No. 9.3e-10;  
Matches 68; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 116 AAGCTCTCTCTCTCCACACACACACACACACACACACACACACACACACACAC 175  
DB 345 AACT 286  
QY 176 ACAACACACACACCCCGC 192  
DB 285 ACACACACACATCACAC 269

## RESULT 15

US-10-027-632-250183  
/ Sequence 250183, Application US/10027632  
/ Publication No. US20030204075A9

## GENERAL INFORMATION:

/ APPLICANT: Wang, David G.  
/ TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
/ TITLE OF INVENTION: Polymorphisms in the Human Genome  
/ FILE REFERENCE: 108827.129  
/ CURRENT APPLICATION NUMBER: US/10/027,632  
/ CURRENT FILING DATE: 2002-04-30  
/ PRIOR APPLICATION NUMBER: US 60/218,006  
/ PRIOR FILING DATE: 2000-07-12  
/ PRIOR APPLICATION NUMBER: US 60/198,676  
/ PRIOR FILING DATE: 2000-04-20  
/ PRIOR APPLICATION NUMBER: US 60/193,483  
/ PRIOR FILING DATE: 2000-03-29  
/ PRIOR APPLICATION NUMBER: US 60/185,218  
/ PRIOR FILING DATE: 2000-02-24  
/ PRIOR APPLICATION NUMBER: US 60/167,363  
/ PRIOR FILING DATE: 1999-11-23  
/ PRIOR APPLICATION NUMBER: US 60/156,358  
/ PRIOR FILING DATE: 1999-09-28  
/ PRIOR APPLICATION NUMBER: US 60/146,002  
/ PRIOR FILING DATE: 1999-08-09  
/ NUMBER OF SEQ ID NOS: 325720  
/ SOFTWARE: FastSeq for Windows Version 4.0  
/ SEQ ID NO 250183  
/ LENGTH: 719  
/ TYPE: DNA  
/ ORGANISM: Human  
US-10-027-632-250183

Query Match 31.7%; Score 63.8; DB 15; Length 719;  
Best Local Similarity 60.8%; Pred. No. 1.2e-09;  
Matches 104; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 24 GTAGAGTAGAGACCAAAATCCACCAAGCTTTATCTATGACACATCCTTTATCT 83  
DB 7 GTAAAGTACTAAGTTAAAAAACTTTAGCCCTCTGTCTCTCCAAAGTTGGTATATTA 66  
QY 84 ATAATTCCAGACACTTTTCTCTCTATGATAGGCTCTCTCTCTCTCTCTCTCTCTCT 143  
DB 67 AATAATGATATCATTTTAAAGCTGGGAGCATGTGAGACTGTGTCTCATATACACACAC 126  
QY 144 ACAAC 194  
DB 127 AC 177

Search completed: March 12, 2004, 09:38:44



QY 44 TATTACTCTTCTTTTTCGAGATGGAGTTTGCTGTG 103



Db	55201	TTTTTATTTGTTTTCTTTTTTTTTTTTTTTTGTAGATGGAGTTGCTCTGTGG	55260
Oy	104	CCCATGCTGGAGTGGAAATGGCATGATCATGCTCATCGAACTCCACCTCTGGGTTTA	163
Db	55261	CCCGAGATGGAGTACCAATGGCAATGATTTTGGCTCGTGCACACTCCACTCCACGGTTTA	55320
Oy	164	AGCAAAAGCTGTGGCTGAGCTCCCGGGTAGATGGATTATACAGCGCCACACCACT	223
Db	55321	AGCATTTCTCTGCTCAGCCTCTCAGATACCTGGGATTACAGTGGCCACCACTATGCC	55380
Oy	224	CGGCTAATGTTTGTATTTTATGTAGTAGAATGGGGTTTACCATGTTGGCCAGGCTGTTC	283
Db	55381	TGGCTAATTTTGTATTTTATGTAGTAGAATGGGGTTTACCATGTGGCCAGGCTGTCTT	55440
Oy	284	AAACTCTGACCTCAG	299
Db	55441	GAACTCTGACCTCAG	55456

```

RESULT 2
US-09-818-512-3/c
; Sequence 3, Application US/09818512
; Patent No. 6537780
; GENERAL INFORMATION:
; APPLICANT: BEASLEY, Ellen et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CLO01192
; CURRENT APPLICATION NUMBER: US/09/818,512
; CURRENT FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 116592
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1) .. (116592)
; OTHER INFORMATION: n = A,T,C or G
US-09-818-512-3

```

```

Query Match      64.5%; Score 194.2; DB 4; Length 116592;
Best Local Similarity   84.4%; Pred. No. 3.7e-37;
Matches 217; Conservative 1; Mismatches 39; Indels 0; Gaps 0;

QY          42 AGATTACTCTTCCTTTTCTTTTTTTTTTTTTTTTTTTTGAGATGGAAGTTTGGCTGT 101
DB          39805 ATTTCGTGAATTTTTTTTTTTTTTTTTTTTTTTTTTTTGTAGATGAGATCACAATCTTGT 39746

QY          102 TGCCCATGCTGAGTGGATGGACATGAYCATAGCTCACTGCAAACCTCACCTCTCGGTT 161
DB          39745 TGCCAGGCTGAGAGGACAGTGGACAACTCAGACACTCAAAGTCCACCTTAGGCT 39688

QY          162 CAAGGAAGCTGTGCGCTCAGCCTCCCGGGAGAATGGAGTTAAGAGGCCCAACCAACA 221
DB          39685 CAAAGCAATTCACAGCCCTCAGCCCTCGAGTAGCTGGAGATTACAGGACCCCACTTAACAA 39628

QY          222 CTGGGCTATGTTTGTATTTTATAGAGAGATGGGGTTTCAACANGTTGGCCAGGCTGTC 281
DB          39625 CCGAGCTAATTTTGTATTTTATAGTAGACAGGGTTTCCGCAATGTTGGCCAGGCTGTC 39566

QY          282 TCAAACTCCTGACCCTCA 298
DB          39565 TCAAACTCCTGACCCTCA 39549

RESULT 3
US-09-791-211-10/c
; Sequence 10, Appraisal US/09791211
; Patent No. 6448080
; Patent INFORMATION:

```

```

APPLICANT: Donna T. Ward
APPLICANT: Andrew T. Walt
TITLE OF INVENTION: ANTISENSE MODULATION OF WRN EXPRESSION
FILE REFERENCE: RTS-0205
CURRENT APPLICATION NUMBER: US/09/791,211
CURRENT FILING DATE: 2001-02-23
NUMBER OF SEQ ID NOS: 90
SEQ ID NO 10
LENGTH: 98844
TYPE: DNA
ORGANISM: Homo sapiens
FEATURES:
NAME/KEY: unsure
LOCATION: 24962
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 64383
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 65468
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 65469
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 65470
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 65471
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 87130
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 89049
OTHER INFORMATION: unknown
OTHER INFORMATION: unknown
US-09-791-211-10
```

```

Query Match      63.7%; Score 191.8; DB 4; Length 98844;
Best Local Similarity 86.1%; Pred. No. 1,3e-36;
Matches 211; Conservative 1; Mismatches 33; Indels 0; Gaps 0;

QY      55 CTTTCTTTTTTTTTTTTTTTTTTTTTTTTGAGATGGAGTGTGTCCTTTGTTGCCCATGCTGA 114
Db       82331 CTTTCTTTTTTTTTTTTTTTTTTTTTTTTGGAGACAGAGTTTGCTCTTTGTTGCCAAGGCTGA 82272

QY      115 GTGGAAATGGCANTGAYCAATAGCTCACTGCACAACCTCCACTCTCGGGTTCAAGCAAAGCTGT 174
Db       82271 GTGGAAATGGCAACATCTGGGCTCACTGCACAACCTCCACTCCGAGTTCAAGCAATTCCCC 82212

QY      175 CGCCTCAGCCCTCCCGGTAGATGGAGTTACAGGCGGCCAACCAACACTCGGCTAATGTT 234
Db       82211 TGCTCTAGCCCTCCCAAGTAGCTGGAGTTACAGGCAACCTCCACCAAGCCAGTAATTGT 82152

QY      235 TGTATTTTGTAGTAGAGATGGGGTTTACCATGTTGGCCAAGCTGTCTCAAACTCCTGAC 294
Db       82151 TGTATTTTGTAGTAGAGACAGGGTTTACCATGTTGACCAAGGCTGTGTTCTGAACCTCCTGAC 82092

QY      295 CTCGAG 299 |||||
Db       82091 CTCGAG 82087 |||||

RESULT 4
US-09-548-797B-7/c
; Sequence 7, Application US/09548797B
; Patent No. 6683165
; GENERAL INFORMATION:
; APPLICANT: KEITH, TIM
; TITLE OF INVENTION: NOVEL HUMAN GENE RELATING TO RESPIRATORY DISEASES AND
; TITLE OF INVENTION: OBESITY
; FILE REFERENCE: 2976-4039

```



```

:
: CURRENT APPLICATION NUMBER: US-09/548,797B
:
: CURRENT FILING DATE: 2002-11-26
:
: PRIOR APPLICATION NUMBER: 60/129,351
:
: PRIOR FILING DATE: 1999-04-13
:
: NUMBER OF SEQ ID NOS: 170
:
: SOFTWARE: PatentIn Ver. 2.1
:
: SEQ ID NO 7
:
: LENGTH: 17000
:
: TYPE: DNA
:
: ORGANISM: Homo sapiens
:
: US-09-548-797E-7

```

	Query Match	62.7%; Score 188.8; DB 4; Length 17000;
	Best Local Similarity	81.6%; Pred. No. 4,33-36;
	Matches 217; Conservative 1; Mismatches 48; Indels 0; Gaps 0;	
OY	35 ATGATCAAGTATTACTCTTTCTTTTTTTTTTTTTTTTTTTTTTTTGAGATGGAGTTTG 94	
Db	16118 ATTAAACGTGTAATTTTAGAGTCCTCTTTTTTTTTTTTTTTTTTTTGAACGGAGCTTG 16055	
OY	95 GTCTTTGTTGCCATGCTGGAGTGSATGCGATGAYCATAGCTA CTGCAA CTTCCACC TC 154	
Db	16058 CTC TTGTGTGCCAGGCTGGAGTGCATGCGACATCTCACTC A CTGCAACCTCCACCTTC 15993	
OY	155 CTGGGTTCAAGCAAGAAGCTGTGCTCGACCTCCCGGTAATGGAGATTACAGCGCCAC 214	
Db	15998 CCATGTTCAACAATTTCTCTCCGCGCGACCTCCCGAATACTGGGATTATAGCAAGTGC 15938	
OY	215 CACCAACTGGGCAATGTTGTATTTTATGTAGATGGGGTTTCAACATGTGTGGCAG 274	
Db	15938 CACCAACACCGGCAATTTTTTGTATTTTATGTAGATGGGGTTTCAACATGTGGCAG 15877	
OY	275 GCTGATCCAACCTCGACCTCAGA 300	
Db	15878 GCTGATCTGAACCTTCGACCTCAAA 15853	

```

RESULT 5
US-09-608-285A-8
Sequence 8, Application US/09608285A
Patent No. 6335013
GENERAL INFORMATION:
APPLICANT: Ford, John
APPLICANT: Mulero, Jullio
APPLICANT: Yeung, George
TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO CD39-LIKE
TITLE OF INVENTION: POLYPEPTIDES
FILE REFERENCE: 28110/36570
CURRENT APPLICATION NUMBER: US/09/608,285A
CURRENT FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: 09/583,231
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 09/557,800
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 09/481,238
PRIOR FILING DATE: 2000-01-11
PRIOR APPLICATION NUMBER: 09/370,265
PRIOR FILING DATE: 1999-08-09
PRIOR APPLICATION NUMBER: PCT/US99/16180
PRIOR FILING DATE: 1999-07-16
PRIOR APPLICATION NUMBER: 09/350,836
PRIOR FILING DATE: 1999-07-09
PRIOR APPLICATION NUMBER: 09/273,447
PRIOR FILING DATE: 1999-03-19
PRIOR APPLICATION NUMBER: 09/244,444
PRIOR FILING DATE: 1999-02-04
PRIOR APPLICATION NUMBER: 09/122,449
PRIOR FILING DATE: 1998-07-24
PRIOR APPLICATION NUMBER: 09/118,205
PRIOR FILING DATE: 1998-07-16
NUMBER OF SEQ ID NOS: 60
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 8

```

```

?      LENGTH: 9365
?      TYPE: DNA
?      ORGANISM: Homo sapiens
?      FEATURE:
?      NAME/KEY: misc_feature
?      LOCATION: (3409)
?      OTHER INFORMATION: n = adenine or guanine or cytosine or thymidine
?      NAME/KEY: misc_feature
?      LOCATION: (9214)
?      OTHER INFORMATION: n = adenine or guanine or cytosine or thymidine
?      NAME/KEY: misc_feature
?      LOCATION: (9303)
?      OTHER INFORMATION: n = adenine or guanine or cytosine or thymidine
?      NAME/KEY: misc_feature
?      LOCATION: (9311)
?      OTHER INFORMATION: n = adenine or guanine or cytosine or thymidine
US-09-608-265A-8

```

[illegible]

```

RESULT 6
US-09/350-8368-8
; Sequence 8, Application US/093508368
; Patent No. 6387645
; GENERAL INFORMATION:
; APPLICANT: Ford, John
; APPLICANT: Mulero, Julio
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL CD39-LIKE
; TITLE OF INVENTION: POLYPEPTIDES
; FILE REFERENCE: 28110/35761
; CURRENT APPLICATION NUMBER: US/09/350.8368
; CURRENT FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: 09/273,447
; PRIOR FILING DATE: 1999-03-19
; PRIOR APPLICATION NUMBER: 09/118,205
; PRIOR FILING DATE: 1998-07-16
; PRIOR APPLICATION NUMBER: 09/122,449
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: 09/244,444
; PRIOR FILING DATE: 1999-02-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 8
; LENGTH: 9365
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: exon
; LOCATION: (1)..(288)
; NAME/KEY: exon

```



```

? RESULT 7
? US-09-370-265-8
? Sequence 8, Application US/09370265
? Patent No. 6447771
? GENERAL INFORMATION:
? APPLICANT: Ford, John
? APPLICANT: Mulero, Jullio
? TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL CD39-LIKE
? TITLE OF INVENTION: POLYPEPTIDES
? FILE REFERENCE: 2811/35908
? CURRENT APPLICATION NUMBER: US/09/370,265
? CURRENT FILING DATE: 1999-08-09
? EARLIER APPLICATION NUMBER: PCT/US99/16180
? EARLIER FILING DATE: 1998-07-16
? EARLIER APPLICATION NUMBER: 09/350,836
? EARLIER FILING DATE: 1999-07-09
? EARLIER APPLICATION NUMBER: 09/273,447
?

```

[illegible]



QY	141	GCACCTCCACCTCCTGGGTTCAAGCAAGCTGTGGCTCAGACCTCCGGGTAAATGGGA	200
QY	7504	GCACCTTCRCCTCTGGTTTAAAGCATTCCTTCCTCAGCCTCCRCAGTAGCTGGGA	7563
QY	201	TTACAGGCGCCACCACCACCACTCGGCTAA--TGTTTATTTTATAGAGAGATGGGGTT	258
QY	7564	TTACAGGCGRGGCGCCACCAAGCGCTGGCTAATTTTTTTATTTTATAGTAGAGACGGGTT	7623
QY	259	TCACCATGTTGGCAGAGCTGGTCTCAAACTCCTGACCTCAG	299
QY	7624	TCACCATGTTGGCAGAGCTRKTCTTACCTCTGATGCTCAG	7664

RESULT 8  
US-09-557-800C-8  
Sequence 8, Application US/09557800C

```

1  APPLICANT: Mulero, Julio
2  APPLICANT: Yeung, George
3  TITLE OF INVENTION: Methods and Materials Relating to CD39-Like
4  TITLE OF INVENTION: Polypeptides
5  FILE REFERENCE: 28110/36457
6  CURRENT APPLICATION NUMBER: US/09/557,800C
7  CURRENT FILING DATE: 2000-04-25
8  PRIOR APPLICATION NUMBER: 09/481,238
9  PRIOR FILING DATE: 2000-01-11
10 PRIOR APPLICATION NUMBER: 09/370,265
11 PRIOR FILING DATE: 1999-08-09
12 PRIOR APPLICATION NUMBER: PCT/US99/16180
13 PRIOR FILING DATE: 1999-07-16
14 PRIOR APPLICATION NUMBER: 09/350836
15 PRIOR FILING DATE: 1999-07-09
16 PRIOR APPLICATION NUMBER: 09/273447
17 PRIOR FILING DATE: 1999-03-19
18 PRIOR APPLICATION NUMBER: 09/122449
19 PRIOR FILING DATE: 1998-07-24
20 PRIOR APPLICATION NUMBER: 09/244444
21 PRIOR FILING DATE: 1998-02-04
22 PRIOR APPLICATION NUMBER: 09/118,205
23 PRIOR FILING DATE: 1998-07-16
24 NUMBER OF SEQ ID NOS: 56
25 SOFTWARE: Patentin Ver. 2.0
26 SEQ ID NO 8
27 LENGTH: 9365
28 TYPE: DNA
29 ORGANISM: Homo sapiens
30 FEATURE:
31 NAME/KEY: misc feature
32 LOCATION: (3405)
33 OTHER INFORMATION: n = adenine or guanine or cytosine or thymidine
34 FEATURE:
35 NAME/KEY: misc_feature
36 LOCATION: (9214)
37 OTHER INFORMATION: n = adenine or guanine or cytosine or thymidine
38 FEATURE:
39 NAME/KEY: misc_feature
40 LOCATION: (9303)
41 OTHER INFORMATION: n = adenine or guanine or cytosine or thymidine
42 FEATURE:
43 NAME/KEY: misc feature
44 LOCATION: (9311)
45 OTHER INFORMATION: n = adenine or guanine or cytosine or thymidine
46
47 US-09-557-800C-8

```

Query Match	62.3%;	Score 187.4;	DB 4;	Length 9365;
Best Local Similarity	75.4%;	Pred. No. 8e-36;		
Matches 212;	Conservative 23;	Mismatches 44;	Indels 2;	Gaps 1;

QY	22	ATTCAAGTTGCACATGATCAGATCAATACCTTCGTTTTTTTTTTTTTTTTTT	80
DB	7384	AATTATAGATGCCAAGTCACATGTTATTATTATTATTATTATTATTATTATT	7443

QY	81	GAGATGGAGATTGGCTCTGTGTGGCCANTGGAGTGGAAATGGACATGACATCTACT	140
Db <td>7444</td> <td>GAGAGGAGATCTGCTCTCTTTTCTTCCTGAGCTGAGTGCARTGGCTRGATCTMGRCCTACT</td> <td>7509</td>	7444	GAGAGGAGATCTGCTCTCTTTTCTTCCTGAGCTGAGTGCARTGGCTRGATCTMGRCCTACT	7509
QY <td>141</td> <td>GCAACCTCCACTCCTCGGTTCAAAGCAAGCTGCGCTCAGCTCCCGGTAGATGGGA</td> <td>200</td>	141	GCAACCTCCACTCCTCGGTTCAAAGCAAGCTGCGCTCAGCTCCCGGTAGATGGGA	200
Db <td>7504</td> <td>GCARCTTCRCCTCTGGGTTCAAGCATTTCTCTGCTCAGCTCCCAAGTAGCTGGGA</td> <td>7563</td>	7504	GCARCTTCRCCTCTGGGTTCAAGCATTTCTCTGCTCAGCTCCCAAGTAGCTGGGA	7563
QY <td>201</td> <td>TTACAGGGCCCAACACACACTGGGTA--TGTTGATTTTGTAGTAGATGGGGT</td> <td>258</td>	201	TTACAGGGCCCAACACACACTGGGTA--TGTTGATTTTGTAGTAGATGGGGT	258
Db <td>7564</td> <td>TTACAGGCATGGGCGACACCGCCGGGTATTTTTTGTATTTTGTAGTAGAAGCGGGTT</td> <td>7623</td>	7564	TTACAGGCATGGGCGACACCGCCGGGTATTTTTTGTATTTTGTAGTAGAAGCGGGTT	7623
QY <td>259</td> <td>TACCATGTGGCGCAGGCTGGTCTCAAACTCTGACTCAG</td> <td>299</td>	259	TACCATGTGGCGCAGGCTGGTCTCAAACTCTGACTCAG	299
Db <td>7624</td> <td>TACCATGTGGCGCAGGCTRKTCTYRACACTCTGAYCTCAG</td> <td>7664</td>	7624	TACCATGTGGCGCAGGCTRKTCTYRACACTCTGAYCTCAG	7664

QY 259 TCACCATGTTGGCCAGGCTGGTCTTCAAACTCTGACCTCAG 299  
 |||||::||::||::||::||::||  
 Db 7624 TCACCATGTTGGCCAGGCTRKTCTYRAACTYCTGAYCTCAG 7664

```

RESULT 9
US-09-370-625A-8
: Sequence 8, Application US/09370625A
: Patent No. 6600032
:
: GENERAL INFORMATION:
:
: APPLICANT: Ford, John
: APPLICANT: Mulero, Julio
:
: TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL CD39-LIKE POLYPEPTIDES
: FILE REFERENCE: 28110/35908
:
: CURRENT APPLICATION NUMBER: US/09/370,625A
:
: CURRENT FILING DATE: 1999-08-09
:
: PRIOR APPLICATION NUMBER: PCT/US99/16180
:
: PRIOR FILING DATE: 1999-07-16
:
: PRIOR APPLICATION NUMBER: 09/350,836
:
: PRIOR FILING DATE: 1999-07-09
:
: PRIOR APPLICATION NUMBER: 09/273,447
:
: PRIOR FILING DATE: 1999-03-19
:
: NUMBER OF SEQ ID NOS: 39
:
: SOFTWARE: PatentIn Ver. 2.0
:
: SEQ ID NO 8
:
: LENGTH: 9365

```

```

1 ORGANISM: Homo sapiens
2 FEATURE:
3 NAME/KEY: exon
4 LOCATION: (1)..(288)
5 FEATURE:
6 NAME/KEY: exon
7 LOCATION: (1281)..(1580)
8 FEATURE:
9 NAME/KEY: exon
10 LOCATION: (1820)..(1855)
11 FEATURE:
12 NAME/KEY: exon
13 LOCATION: (12467)..(2555)
14 NAME/KEY: exon
15 LOCATION: (3889)..(3950)
16 NAME/KEY: exon
17 LOCATION: (4894)..(4995)
18 NAME/KEY: exon
19 LOCATION: (5847)..(5987)
20 NAME/KEY: exon
21 LOCATION: (6966)..(7138)
22 NAME/KEY: exon
23 LOCATION: (8566)..(9365)
24 NAME/KEY: misc_feature

```



```

? LOCATION: (3409)
? OTHER_INFORMATION: n = a or g or t or c
?
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (9214)
? OTHER_INFORMATION: n = a or g or t or c
?
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (9303)
? OTHER_INFORMATION: n = a or g or t or c
?
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (9311)
? OTHER_INFORMATION: n = a or g or t or c
?
US-09-370-625A-8

```

Query Match	62.3%;	Score 187.4;	DB 4;	Length 9365;
Best Local Similarity	75.4%;	Pred. No. 8e-36;		
Matches 212;	Conservative 23;	Mismatches 44;	Indels 2;	Gaps 1.

OY	21	ATTCAAGTTGCATGACGTGAACAAGTAACTCTTCCTTTTTTTTTTTTTTTTTTT	80
Dd	7364	AATTAGATGCCAAGTACATCGTTATTATTATTATTAATTATTATTATTATT	7443
OY	81	GAGATGAGTTTGCTCTTGTTCCCAAGCTGGAGTGGATGGCATGAYCATAGTCACT	140
Dd	7444	GAGAAGGAGTCTYGCTCTTGTTCCATGGCTGGAGTGCARTGGCAGTACWCRGGTCACT	7503
OY	141	GGAACCTCACACTCTCGGSGTTCAAGGAAGCGTGCGCCACAGCTCCCGGATAGATGGGA	200
Dd	7504	GGARCTTCCTCTCTGSGTTCAAAGGATTTCTCTGCTCAGGCTCCCAAGTAGCTGGSA	7563
OY	201	TTACAGCGCCCCACCAACAACACTCGGCTAA--TGTTTGATTTTTAGTAGAGATGGGTT	258
Dd	7564	TACAGGCGRGCGCCAACAAGCCGCGGCTAATTTTTTTGATTTTTAGTAGAGACGGGGTT	7623
OY	259	TCACCATGTTGGCCAGGCTGCTCAATACTCGTACCTAG	299
Dd	7624	TCACCATGTTGGCCAGGCTRKTCTTAACTCTGATCTAG	7664

```

RESULT 10
; Sequence 42, Application US/09608285A
; Patent No. 6335013
; GENERAL INFORMATION:
; APPLICANT: Ford, John
; APPLICANT: Mulero, Julio
; APPLICANT: Yeung, George
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO CD39-LIKE
; FILE REFERENCE: 28110/36570
; CURRENT APPLICATION NUMBER: US/09/608,285A
; CURRENT FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 09/583,231
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 09/557,800
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/481,238
; PRIOR FILING DATE: 2000-01-11
; PRIOR APPLICATION NUMBER: 09/370,265
; PRIOR FILING DATE: 1999-08-09
; PRIOR APPLICATION NUMBER: PCT/US99/16180
; PRIOR FILING DATE: 1999-07-16
; PRIOR APPLICATION NUMBER: 09/350,836
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: 09/273,447
; PRIOR FILING DATE: 1999-03-19
; PRIOR APPLICATION NUMBER: 09/244,444
; PRIOR FILING DATE: 1999-02-04
; PRIOR APPLICATION NUMBER: 09/122,449
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: 09/118,205

```

```

; PRIOR FILING DATE: 1998-07-16
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 42
; LENGTH: 14747
; TYPE: DNA
; ORGANISM: Homo sapiens
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (13641)
; OTHER INFORMATION: n = adenosine or guanine or cytosine or thymidine
US-09-608-285A-42

Query Match          62.3%; Score 187.4; DB 4; Length 14747;
Best Local Similarity 75.4%; Pred. No. 9e-36;
Matches 212; Conservative 23; Mismatches 44; Indels 2; Gaps 1.

```

[illegible]

```

1  RESULT 11
2  US-09-557-800C-42
3  ; Sequence 42, Application US/09557800C
4  ; Patent No. 6476211
5  ; GENERAL INFORMATION:
6  ; APPLICANT: Ford, John
7  ; APPLICANT: Malero, Julio
8  ; APPLICANT: Young, George
9  ; TITLE OF INVENTION: Methods and Materials Relating to CD39-Like
10 ; TITLE OF INVENTION: Polypeptides
11 ; FILE REFERENCE: 28110/36457
12 ; CURRENT APPLICATION NUMBER: US/09/557,800C
13 ; CURRENT FILING DATE: 2000-04-25
14 ; PRIOR APPLICATION NUMBER: 09/461,238
15 ; PRIOR FILING DATE: 2000-01-11
16 ; PRIOR APPLICATION NUMBER: 09/370,265
17 ; PRIOR FILING DATE: 1999-08-09
18 ; PRIOR APPLICATION NUMBER: PCT/US99/16180
19 ; PRIOR FILING DATE: 1999-07-16
20 ; PRIOR APPLICATION NUMBER: 09/350836
21 ; PRIOR FILING DATE: 1999-07-09
22 ; PRIOR APPLICATION NUMBER: 09/273447
23 ; PRIOR FILING DATE: 1999-03-19
24 ; PRIOR APPLICATION NUMBER: 09/122449
25 ; PRIOR FILING DATE: 1998-07-24
26 ; PRIOR APPLICATION NUMBER: 09/244444
27 ; PRIOR FILING DATE: 1999-02-04
28 ; PRIOR APPLICATION NUMBER: 09/118,205
29 ; PRIOR FILING DATE: 1998-07-16
30 ; NUMBER OF SEQ ID NOS: 56
31 ; SOFTWARE: PatentIn Ver. 2.0
32 ; SEQ ID NO 42
33 ; LENGTH: 14747
34 ; TYPE: DNA
35 ; ORGANISM: Homo sapiens

```











GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 11, 2004, 17:19:49 ; Search time 528.064 Seconds  
(without alignments)  
2098.537 Million cell updates/sec

Title: US-09-966-880A-9\_COPY\_5214\_5514  
Perfect score: 301  
Sequence: 1 cttctctcttttccatagtt.....tcaactcctgactcagag 301

Scoring table:  
IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 2432557 seqs, 1840798884 residues

Total number of hits satisfying chosen parameters: 4865114

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA:\*

- 1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/2/pubpna/PT1\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/2/pubpna/PTCT5\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*
- 14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq:\*
- 16: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*
- 17: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*
- 18: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	300.6	99.9	5514	9	US-09-966-880A-9
2	300.6	99.9	11204	9	US-09-966-880A-35
C 3	204	67.8	26657	9	US-09-810-673A-3
C 4	204	67.8	26657	14	US-10-395-242-3
5	199.2	66.2	626	15	US-10-027-632-133310
6	199	66.1	177556	10	US-09-952-213D-6
7	198.8	66.0	626	15	US-10-027-632-133311
8	196.4	65.2	15173	9	US-09-764-869-1981
9	196.4	65.2	15173	14	US-10-091-504-1981
10	196.4	65.2	15173	15	US-10-227-577-1981
11	196.4	65.2	42595	11	US-09-984-429-469
12	196	65.1	2460	15	US-10-108-260A-595
C 13	195.8	65.0	777	15	US-10-027-632-172403
C 14	195	64.8	557	15	US-10-027-632-135991
C 15	194.8	64.7	6254	9	US-09-764-868-1495

C 16	194.2	64.5	116592	9	US-09-818-512-3	Sequence 3, Appl1
C 17	194.2	64.5	116592	14	US-10-354-065-3	Sequence 3, Appl1
18	194	64.5	164875	15	US-10-085-117-322	Sequence 332, App
19	193.6	64.3	2136	15	US-10-027-632-110040	Sequence 110040,
C 20	193.2	64.2	9556	15	US-10-027-632-250166	Sequence 250166,
21	193.2	64.2	9556	10	US-09-764-872-481	Sequence 481, App
22	193.2	64.2	9557	10	US-09-764-872-479	Sequence 479, App
23	193	64.1	145831	9	US-09-969-708-79	Sequence 79, Appl
24	193	64.1	145831	9	US-09-954-455-2116	Sequence 2116, Ap
25	193	64.1	145831	10	US-09-873-357C-646	Sequence 646, App
26	193	64.1	145831	12	US-10-240-425-363	Sequence 363, App
C 27	193	64.1	358246	15	US-10-292-798-1095	Sequence 1095, Ap
C 28	192.6	64.0	5781	9	US-09-764-868-1374	Sequence 1374, Ap
C 29	192.6	64.0	5782	9	US-09-764-868-1375	Sequence 1375, Ap
C 30	192.4	63.9	23934	9	US-09-764-860-777	Sequence 777, App
C 31	192.4	63.9	23934	9	US-09-764-877-2536	Sequence 2536, App
C 32	192.4	63.9	23934	9	US-09-764-877-2544	Sequence 2544, Ap
C 33	192.4	63.9	23934	10	US-09-764-891-7210	Sequence 7210, App
C 34	192.4	63.9	23934	14	US-10-074-095-777	Sequence 777, App
C 35	192.4	63.9	23934	15	US-10-212-872-717	Sequence 717, App
C 36	192.4	63.9	23934	15	US-10-242-515-2536	Sequence 2536, App
C 37	192.4	63.9	23934	15	US-10-242-515-2544	Sequence 2544, Ap
C 38	192	63.8	681	15	US-10-027-632-137222	Sequence 137222,
C 39	192	63.8	23130	9	US-09-764-869-1892	Sequence 1892, Ap
C 40	192	63.8	23130	14	US-10-091-504-1892	Sequence 1892, Ap
C 41	192	63.8	23130	15	US-10-227-577-1892	Sequence 1892, Ap
C 42	191.4	63.6	5797	10	US-09-764-891-6093	Sequence 6093, App
C 43	191.2	63.5	1035	14	US-10-196-846-4227	Sequence 4227, App
C 44	191	63.5	3150	9	US-09-764-870-622	Sequence 622, App
C 45	191	63.5	3150	14	US-10-125-540-622	Sequence 622, App

## ALIGNMENTS

RESULT 1  
US-09-966-880A-9  
Sequence 9, Application US/09966880A  
Patent No. US20020164743A1  
GENERAL INFORMATION:  
APPLICANT: Honjo, Tasuku  
TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
FILE REFERENCE: 06501-088001  
CURRENT APPLICATION NUMBER: US/09/966,880A  
CURRENT FILING DATE: 2001-09-28  
PRIOR APPLICATION NUMBER: PCT/JP00/01918  
PRIOR FILING DATE: 2000-03-28  
PRIOR APPLICATION NUMBER: JP 11-371382  
PRIOR FILING DATE: 1999-12-27  
PRIOR APPLICATION NUMBER: JP 11-178999  
PRIOR FILING DATE: 1999-06-24  
PRIOR APPLICATION NUMBER: JP 11-87192  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 9  
LENGTH: 5514  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: Intron  
LOCATION: (1)...(1031)  
FEATURE:  
NAME/KEY: exon  
LOCATION: (1032)...(1118)  
FEATURE:  
NAME/KEY: Intron  
LOCATION: (1119)...(5514)  
US-09-966-880A-9  
Query Match 99.9%; Score 300.6; DB 9; Length 5514;  
Best Local Similarity 100.0%; Pred. No. 5.8e-61;



	Matches	301;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	1	CTCTTCAATTTTCCATAGATATCAAGTTTGACATATGATCAAGATATTTCTCTTCTTTT							60	
Db	5214	CTCTTCATATTTTCCATAGTATTTCAAGTTTACACATGATCAAGATATTTCTTTCTTTT							5273	
QY	61	TTTTTTTTTTTTTTTTTTTGAAGATGAGATTTGGTCTGTGTGCCCATGCTGAGTGGAA							120	
Db	5274	TTTTTTTTTTTTTTTTTTTGAAGATGAGATTTGGTCTGTGTGCCCATGCTGAGTGGAA							5333	
QY	121	TGGCATGATCATAGCTCACTGCAACCTCCACTCTGGGTTCAAGCAAAAGCTGTGCCTC							180	
Db	5334	TGGCATGATCATAGCTCACTGCAACCTCCACTCTGGGTTCAAGCAAAAGCTGTGCCTC							5393	
QY	181	AGCCTCCCGGGTAGATGAGATTTACAGGCGGCCACACACACACTCGGCTTAATGTTTGAAT							240	
Db	5394	AGCCTCCCGGGTAGATGAGATTTACAGGCGGCCACACACACACTCGGCTTAATGTTTGAAT							5453	
QY	241	TTTAAATAGATGGGGTTTCAACATGTTGGCCAGGCTGTGTCAAACTCTGACCTCAGA							300	
Db	5454	TTTAAATAGATGGGGTTTCAACATGTTGGCCAGGCTGTGTCAAACTCTGACCTCAGA							5513	
QY	301	G 301								
Db	5514	G 5514								

```

RESULT 2
US-09-966-880A-35
; Sequence 35, Application US/09966880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; APPLICANT: Muramatsu, Masamichi
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966, 880A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 35
; LENGTH: 11204
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-966-880A-35

```

Query Match	99.98%	Score 300.6;	DB 9;	Length 11204;
Best Local Similarity	99.78%	Pred. No. 7.4e-61;		
Matches 300; Conservative	1;	Mismatches 0;	Indels 0;	Gaps 0

[illegible]

QY 241 TTTAGTAGAGATGGGGTTTGCACATGTTGGGCCAGGCTGGTCTCAAATCCTGACCTGACA 300  
 Db 4864 TTTAGTAGAGATGGGGTTTGCACATGTTGGGCCAGGCTGGTCTCAAATCCTGACCTGACA 4924  
 QY 301 G 301  
 Db 4924 G 4924

```

US-09-810-673A-3
RESULT 3
US-09-810-673A-3/c
; Sequence 3, Application US/09810673A
; Patent No. US20020064821A1
; GENERAL INFORMATION:
; APPLICANT: MERRILOV, Gennady et al.
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CLO00723
; CURRENT APPLICATION NUMBER: US/09/810,673A
; CURRENT FILING DATE: 2001-03-19
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 26657
; TYPE: DNA
; ORGANISM: Human
US-09-810-673A-3

```

Query Match	67.8%	Score 204	DB 9	Length 26657
Best Local Similarity	87.4%	Pred. No. 5.4e-38		
Matches 222	Conservative	1	Mismatches 31	Indels 0
				Gaps 0
Qy	44	TATACCTCTCTTTT	TTTTTTTTTTTTTTTTTTTTTTT	TAGATGAGTTTGGTCTTGTG 1033
Db	445	TTTTTTTTTTTTTTTT	TTTTTTTTTTTTTTTTTTTTTTTTTTT	TGAGATGAGTTTGGCTTTATG 386
Qy	104	CCCATGCTGAGTGGATGGATGGATGAYCATAGCTCACTGCACCTTCACCTCTGGGTTCA 1533		
Db	385	CCGAGGCTGAGTGGATGGATGGACAGATCTCAGTCCACCAACTCCGCTCCCGGGTTCA 326		
Qy	164	AGCAAGCTGTGGCTTCAGCTCCGGGATGATGGATTAAGGCGCCACACACACAAT 2233		
Db	325	AGCATGCTCCCGCCTCAGGCTCCCAAGTAGCTGAGATTACAGGGGCACACACACAGCC 266		
Qy	224	CGGCTATGTTTGTATTTT	TAGAGAGATGGGGGTTTCA	ACATGTTGGCAGAGCTGATCTC 2833
Db	265	CGGCTAATTTTTTGTATTTT	TAGTAGAGAGCGGGGTTTCA	ACCATGTTGGTAAAGCTGATCTC 206
Qy	284	AAACTCCGACCTC 297		
Db	205	AAACTCCGACCTC 192		

```

/ RESULT 4
/ US-10-395-242-3/C
/ Sequence 3, Application US/10395242
/ Publication No. US20030166154A1
/
/ GENERAL INFORMATION:
/
/ APPLICANT: MERKULOV, Gennady et al.
/
/ TITLE OF INVENTION: ISOLATED HUMAN ION CHANNEL PROTEINS,
/
/ TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN ION CHANNEL PROTEINS
/
/ TITLE OF INVENTION: AND USES THEREOF
/
/ FILE REFERENCE: CLO00723CON
/
/ CURRENT APPLICATION NUMBER: US/10/395,242
/
/ CURRENT FILING DATE: 2003-03-25
/
/ NUMBER OF SEQ ID NOS: 7
/
/ SOFTWARE: FastSeq for Windows Version 4.0
/
/ SEQ ID NO 3
/
/ LENGTH: 26657
/
/ TYPE: DNA
/
/ ORGANISM: Human
/

```



















Mon Mar 15 09:28:04 2004

us-09-966-880a-9\_copy\_5214\_5514.rnpb

**Page 7**

```

: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: SITE
: LOCATION: (2676)
: OTHER INFORMATION: n equals a,t,g, or c
US-09-764-868-1455

```

Query Match	64.7%	Score 194.8;	DB 9;	Length 6254;
Best Local Similarity	83.3%;	Pred. No. 4.9e-36;		
Matches 220; Conservative	1;	Mismatches 43;	Indels 0;	Gaps 0;

QY	36	TGATCAGATATACCTTTCTTTT	TTTTTTTTTTTTTTTTTTTTTTTTTTTGAATGAGATTGG	95
Db	4109	TGAGCAGATGTCCTGGCCTTTT	TTTTTTTTTTTTTTTTTTTTTTTTTTTGAATGAGATTGG	4055
QY	96	TCTTGTGCCCATGCTGGAGTGG	AATGCGATGATCACTGCACCTCCACCTCC	155
Db	4049	TCTTGTGCCCATGCTGGAGTGG	AATGCGATGATCACTGCACCTCCACCTCC	3995
QY	156	TGGGTTCAAGCAAACTGTGCGCT	CAGGCTCCCGGCTGATGGGATTACAGGCGCCACG	215
Db	3995	CAGGTTCTAGTAACTTCTCTGCT	TACGCTCCCGAGTGGCTGGGATTACAGGCAACAAC	3935
QY	216	ACCACACTCGGCTAAATGTTGAT	TTTTTTAGTAGAGATGGGTTTCAACATGTTGGCAG	275
Db	3929	ACCACACCAAGTTAATTTTGGAT	TTTTTTAGTAGAGATGGGTTTGCACATGTTGGCAG	3875
QY	276	CTGGTCTCAACTCCTGACCTCAG	299	
Db	3869	CTGGTCTCAACTCCTGACCTCAG	3846	

Search completed: March 12, 2004, 09:38:46  
Job time : 530.064 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: March 11, 2004, 17:19:44 ; Search time 2439.15 Seconds

(Without alignments)  
4342.094 Million cell updates/sec

Title: US-09-966-880A-9\_COPY\_5214\_5514

Perfect score: 301

Sequence: 1 cctctctatttctcatagt.....tcaactcctgcaccctcagag 301

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 37577330 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending Patents NA Main:\*

1: /cgn2\_6/ptodata/2/pna/PCRNUS\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/pna/US06\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/pna/US07\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/pna/US080\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/pna/US081\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/pna/US082\_COMB.seq.\*  
7: /cgn2\_6/ptodata/2/pna/US083\_COMB.seq.\*  
8: /cgn2\_6/ptodata/2/pna/US084\_COMB.seq.\*  
9: /cgn2\_6/ptodata/2/pna/US085\_COMB.seq.\*  
10: /cgn2\_6/ptodata/2/pna/US086\_COMB.seq.\*  
11: /cgn2\_6/ptodata/2/pna/US087\_COMB.seq.\*  
12: /cgn2\_6/ptodata/2/pna/US088\_COMB.seq.\*  
13: /cgn2\_6/ptodata/2/pna/US089\_COMB.seq.\*  
14: /cgn2\_6/ptodata/2/pna/US090\_COMB.seq.\*  
15: /cgn2\_6/ptodata/2/pna/US091\_COMB.seq.\*  
16: /cgn2\_6/ptodata/2/pna/US092A\_COMB.seq.\*  
17: /cgn2\_6/ptodata/2/pna/US092B\_COMB.seq.\*  
18: /cgn2\_6/ptodata/2/pna/US093A\_COMB.seq.\*  
19: /cgn2\_6/ptodata/2/pna/US093B\_COMB.seq.\*  
20: /cgn2\_6/ptodata/2/pna/US094\_COMB.seq.\*  
21: /cgn2\_6/ptodata/2/pna/US095A\_COMB.seq.\*  
22: /cgn2\_6/ptodata/2/pna/US095B\_COMB.seq.\*  
23: /cgn2\_6/ptodata/2/pna/US095C\_COMB.seq.\*  
24: /cgn2\_6/ptodata/2/pna/US095D\_COMB.seq.\*  
25: /cgn2\_6/ptodata/2/pna/US095A\_COMB.seq.\*  
26: /cgn2\_6/ptodata/2/pna/US096B\_COMB.seq.\*  
27: /cgn2\_6/ptodata/2/pna/US096C\_COMB.seq.\*  
28: /cgn2\_6/ptodata/2/pna/US096D\_COMB.seq.\*  
29: /cgn2\_6/ptodata/2/pna/US096E\_COMB.seq.\*  
30: /cgn2\_6/ptodata/2/pna/US097A\_COMB.seq.\*  
31: /cgn2\_6/ptodata/2/pna/US097B\_COMB.seq.\*  
32: /cgn2\_6/ptodata/2/pna/US097C\_COMB.seq.\*  
33: /cgn2\_6/ptodata/2/pna/US098A\_COMB.seq.\*  
34: /cgn2\_6/ptodata/2/pna/US098B\_COMB.seq.\*  
35: /cgn2\_6/ptodata/2/pna/US098C\_COMB.seq.\*  
36: /cgn2\_6/ptodata/2/pna/US098D\_COMB.seq.\*  
37: /cgn2\_6/ptodata/2/pna/US099A\_COMB.seq.\*  
38: /cgn2\_6/ptodata/2/pna/US099B\_COMB.seq.\*  
39: /cgn2\_6/ptodata/2/pna/US099C\_COMB.seq.\*  
40: /cgn2\_6/ptodata/2/pna/US099D\_COMB.seq.\*  
41: /cgn2\_6/ptodata/2/pna/US099E\_COMB.seq.\*  
42: /cgn2\_6/ptodata/2/pna/US099F\_COMB.seq.\*  
43: /cgn2\_6/ptodata/2/pna/US100A\_COMB.seq.\*

44: /cgn2\_6/ptodata/2/pna/US100B\_COMB.seq.\*  
45: /cgn2\_6/ptodata/2/pna/US101A\_COMB.seq.\*  
46: /cgn2\_6/ptodata/2/pna/US101B\_COMB.seq.\*  
47: /cgn2\_6/ptodata/2/pna/US102A\_COMB.seq.\*  
48: /cgn2\_6/ptodata/2/pna/US102B\_COMB.seq.\*  
49: /cgn2\_6/ptodata/2/pna/US103A\_COMB.seq.\*  
50: /cgn2\_6/ptodata/2/pna/US103B\_COMB.seq.\*  
51: /cgn2\_6/ptodata/2/pna/US104A\_COMB.seq.\*  
52: /cgn2\_6/ptodata/2/pna/US104B\_COMB.seq.\*  
53: /cgn2\_6/ptodata/2/pna/US106\_COMB.seq.\*  
54: /cgn2\_6/ptodata/2/pna/US107A\_COMB.seq.\*  
55: /cgn2\_6/ptodata/2/pna/US107B\_COMB.seq.\*  
56: /cgn2\_6/ptodata/2/pna/US6000\_COMB.seq.\*  
57: /cgn2\_6/ptodata/2/pna/US6001\_COMB.seq.\*  
58: /cgn2\_6/ptodata/2/pna/US6002\_COMB.seq.\*  
59: /cgn2\_6/ptodata/2/pna/US6003\_COMB.seq.\*  
60: /cgn2\_6/ptodata/2/pna/US6004\_COMB.seq.\*  
61: /cgn2\_6/ptodata/2/pna/US6005\_COMB.seq.\*  
62: /cgn2\_6/ptodata/2/pna/US6006\_COMB.seq.\*  
63: /cgn2\_6/ptodata/2/pna/US6007\_COMB.seq.\*  
64: /cgn2\_6/ptodata/2/pna/US6008\_COMB.seq.\*  
65: /cgn2\_6/ptodata/2/pna/US6009\_COMB.seq.\*  
66: /cgn2\_6/ptodata/2/pna/US6010\_COMB.seq.\*  
67: /cgn2\_6/ptodata/2/pna/US6011\_COMB.seq.\*  
68: /cgn2\_6/ptodata/2/pna/US6012\_COMB.seq.\*  
69: /cgn2\_6/ptodata/2/pna/US6013\_COMB.seq.\*  
70: /cgn2\_6/ptodata/2/pna/US6014\_COMB.seq.\*  
71: /cgn2\_6/ptodata/2/pna/US6015\_COMB.seq.\*  
72: /cgn2\_6/ptodata/2/pna/US6016\_COMB.seq.\*  
73: /cgn2\_6/ptodata/2/pna/US6017\_COMB.seq.\*  
74: /cgn2\_6/ptodata/2/pna/US6018\_COMB.seq.\*  
75: /cgn2\_6/ptodata/2/pna/US6019\_COMB.seq.\*  
76: /cgn2\_6/ptodata/2/pna/US6020\_COMB.seq.\*  
77: /cgn2\_6/ptodata/2/pna/US6021\_COMB.seq.\*  
78: /cgn2\_6/ptodata/2/pna/US6022\_COMB.seq.\*  
79: /cgn2\_6/ptodata/2/pna/US6023A\_COMB.seq.\*  
80: /cgn2\_6/ptodata/2/pna/US6023B\_COMB.seq.\*  
81: /cgn2\_6/ptodata/2/pna/US6024\_COMB.seq.\*  
82: /cgn2\_6/ptodata/2/pna/US6025\_COMB.seq.\*  
83: /cgn2\_6/ptodata/2/pna/US6026\_COMB.seq.\*  
84: /cgn2\_6/ptodata/2/pna/US6027\_COMB.seq.\*  
85: /cgn2\_6/ptodata/2/pna/US6028\_COMB.seq.\*  
86: /cgn2\_6/ptodata/2/pna/US6029\_COMB.seq.\*  
87: /cgn2\_6/ptodata/2/pna/US6030\_COMB.seq.\*  
88: /cgn2\_6/ptodata/2/pna/US6031\_COMB.seq.\*  
89: /cgn2\_6/ptodata/2/pna/US6032\_COMB.seq.\*  
90: /cgn2\_6/ptodata/2/pna/US6033\_COMB.seq.\*  
91: /cgn2\_6/ptodata/2/pna/US6034\_COMB.seq.\*  
92: /cgn2\_6/ptodata/2/pna/US6035\_COMB.seq.\*  
93: /cgn2\_6/ptodata/2/pna/US6036\_COMB.seq.\*  
94: /cgn2\_6/ptodata/2/pna/US6037\_COMB.seq.\*  
95: /cgn2\_6/ptodata/2/pna/US6038\_COMB.seq.\*  
96: /cgn2\_6/ptodata/2/pna/US6039\_COMB.seq.\*  
97: /cgn2\_6/ptodata/2/pna/US6040\_COMB.seq.\*  
98: /cgn2\_6/ptodata/2/pna/US6041\_COMB.seq.\*  
99: /cgn2\_6/ptodata/2/pna/US6042\_COMB.seq.\*  
100: /cgn2\_6/ptodata/2/pna/US6043\_COMB.seq.\*  
101: /cgn2\_6/ptodata/2/pna/US6044\_COMB.seq.\*  
102: /cgn2\_6/ptodata/2/pna/US6045\_COMB.seq.\*  
103: /cgn2\_6/ptodata/2/pna/US6046\_COMB.seq.\*  
104: /cgn2\_6/ptodata/2/pna/US6047\_COMB.seq.\*  
105: /cgn2\_6/ptodata/2/pna/US6048\_COMB.seq.\*  
106: /cgn2\_6/ptodata/2/pna/US6049\_COMB.seq.\*  
107: /cgn2\_6/ptodata/2/pna/US6050\_COMB.seq.\*  
108: /cgn2\_6/ptodata/2/pna/US6051\_COMB.seq.\*  
109: /cgn2\_6/ptodata/2/pna/US6052\_COMB.seq.\*  
110: /cgn2\_6/ptodata/2/pna/US6053\_COMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES



Result No.	Score	Query Match	Length	DB	ID	Description
1	300.6	99.9	5514	41	US-09-966-880A-9	Sequence 9, Appli
2	300.6	99.9	11204	1	PCT-US02-06518-201	Sequence 201, App
3	300.6	99.9	11204	1	US-09-966-880A-35	Sequence 35, Appl
4	300.6	99.9	14684	39	US-09-949-016-13942	Sequence 13942, A
5	299.8	99.9	155313	77	US-60-212-664-16	Sequence 16, Appl
6	299.8	99.6	22683	103	US-60-465-241-52403	Sequence 52403, A
7	299.8	99.6	22683	104	US-60-470-166-8933	Sequence 8933, Ap
8	208.8	69.4	3037287	39	US-09-947-911-286	Sequence 286, App
9	204	67.8	26657	1	PCT-US01-21148-3	Sequence 3, Appli
10	204	67.8	26657	33	US-09-810-673A-3	Sequence 3, Appli
11	204	67.8	26657	50	US-10-395-242-3	Sequence 3, Appli
12	204	67.8	32768	77	US-60-213-177-315	Sequence 315, App
13	204	67.8	32768	77	US-60-213-178-330	Sequence 330, App
14	204	67.8	32768	77	US-60-213-178-330	Sequence 16, Appl
15	202.8	67.4	32768	76	US-60-206-600-43	Sequence 43, Appl
16	202.8	67.4	32768	76	US-60-208-965-61	Sequence 61, Appl
17	202.8	67.4	32768	76	US-60-208-965-61	Sequence 17, App
18	202.8	67.4	32768	77	US-60-213-178-174	Sequence 17, App
19	202.8	67.4	42086	103	US-60-465-241-52589	Sequence 52589, A
20	202.8	67.4	42086	103	US-60-465-412-86320	Sequence 86320, A
21	202.8	67.4	42086	105	US-60-487-610-19732	Sequence 19732, A
22	202.8	67.4	888565	39	US-09-947-916-97	Sequence 97, Appl
23	201.4	66.9	439571	39	US-09-948-128-209	Sequence 209, App
24	201.4	66.9	4057573	39	US-09-948-128-219	Sequence 219, App
25	201.2	66.8	85668	103	US-60-466-412-86802	Sequence 86802, A
26	201.2	66.8	93335	77	US-60-212-664-313	Sequence 313, App
27	201	66.8	44342	39	US-09-949-016-12661	Sequence 12661, A
28	200.6	66.6	16907	77	US-60-230-435-905	Sequence 905, App
29	200.6	66.6	32768	77	US-60-213-172-33	Sequence 33, Appl
30	200.6	66.6	32768	77	US-60-213-173-33	Sequence 33, Appl
31	200	66.4	17677	1	PCT-US01-01354-25318	Sequence 25318, A
32	200	66.4	17677	1	PCT-US01-01354-38907	Sequence 38907, A
33	200	66.4	17677	32	US-09-760-443-1760	Sequence 1760, Ap
34	200	66.4	17677	32	US-09-764-905-25318	Sequence 25318, A
35	200	66.4	17677	32	US-09-764-905-38907	Sequence 38907, A
36	200	66.4	17677	44	US-10-092-399-25318	Sequence 25318, A
37	200	66.4	17677	44	US-10-092-399-38907	Sequence 38907, A
38	200	66.4	17677	47	US-10-212-054-1760	Sequence 1760, Ap
39	200	66.4	41481	107	US-60-500-337-19363	Sequence 19363, A
40	199.6	66.3	41863	39	US-09-949-016-14948	Sequence 14948, A
41	199.4	66.2	392128	103	US-60-465-241-51812	Sequence 51812, A
42	199.4	66.2	888565	39	US-09-947-916-97	Sequence 97, Appl
43	199.2	66.2	626	26	US-09-634-3068-133310	Sequence 133310, A
44	199.2	66.2	626	43	US-10-027-632-133310	Sequence 133310, A
45	199.2	66.2	32768	76	US-60-208-973-84	Sequence 84, Appl

## ALIGNMENTS

RESULT 1  
 US-09-966-880A-9  
 ; Sequence 9, Application US/09966880A  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Honjo, Tasuku  
 ; APPLICANT: Mutamatsu, Masamichi  
 ; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
 ; FILE REFERENCE: 06501-088001  
 ; CURRENT APPLICATION NUMBER: US/09/966, 880A  
 ; CURRENT FILING DATE: 2001-09-28  
 ; PRIOR APPLICATION NUMBER: PCT/JP00/01918  
 ; PRIOR FILING DATE: 2000-03-28  
 ; PRIOR APPLICATION NUMBER: JP 11-371382  
 ; PRIOR FILING DATE: 1999-12-27  
 ; PRIOR APPLICATION NUMBER: JP 11-178999  
 ; PRIOR FILING DATE: 1999-06-24  
 ; PRIOR APPLICATION NUMBER: JP 11-87192  
 ; PRIOR FILING DATE: 1999-03-29  
 ; NUMBER OF SEQ ID NOS: 36

SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 9  
 ; LENGTH: 5514  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: Intron  
 ; LOCATION: (1)...(1031)  
 ; FEATURE:  
 ; NAME/KEY: exon  
 ; LOCATION: (1032)...(1118)  
 ; FEATURE:  
 ; NAME/KEY: Intron  
 ; LOCATION: (1119)...(5514)  
 ; US-09-966-880A-9

Query Match 99.9%; Score 300.6; DB 41; Length 5514;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-25;  
 Matches 301; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 CTTCTATTTTTCATAGTATTCAGTTTGACATGATCAAGTATTACTCTTTCTTTT 60  
 5214 CTCTCTATTTTTCATAGTATTCAGTTTGACATGATCAAGTATTACTCTTTT 5273  
 61 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 120  
 5274 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 5333  
 121 TGGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180  
 5334 TGGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5393  
 131 AGCTCCCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240  
 5394 AGCTCCCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5453  
 241 TTTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300  
 5454 TTTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5513  
 301 G 301  
 5514 G 5514

RESULT 2  
 PCT-US02-06518-201  
 ; Sequence 201, Application PC/TUS0206518  
 ; GENERAL INFORMATION:  
 ; APPLICANT: CONFORMA THERAPEUTICS CORP.  
 ; TITLE OF INVENTION: METHODS FOR TREATING GENETICALLY-DEFINED PROLIFERATIVE  
 ; DISORDERS WITH HSP90 INHIBITORS  
 ; FILE REFERENCE: 031164.001000  
 ; CURRENT APPLICATION NUMBER: PCT/US02/06518  
 ; CURRENT FILING DATE: 2002-03-01  
 ; PRIOR APPLICATION NUMBER: 60/272,751  
 ; PRIOR FILING DATE: 2001-03-01  
 ; NUMBER OF SEQ ID NOS: 330  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 201  
 ; LENGTH: 11204  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; PCT-US02-06518-201

Query Match 99.9%; Score 300.6; DB 1; Length 11204;  
 Best Local Similarity 99.7%; Pred. No. 1.5e-25;  
 Matches 300; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

1 CTTCTATTTTTCATAGTATTCAGTTTGACATGATCAAGTATTACTCTTTCTTTT 60  
 4624 CTCTCTATTTTTCATAGTATTCAGTTTGACATGATCAAGTATTACTCTTTCTTTT 4683







SEQ ID NO 16  
LENGTH: 155313  
TYPE: DNA  
ORGANISM: HUMAN  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)-(155313)  
OTHER INFORMATION: n = A,T,C or G  
US-60-212-664-16

Query Match 99.9%; Score 300.6; DB 77; Length 155313;  
Best Local Similarity 99.7%; Pred. No. 9e-26;  
Matches 300; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCTCTATTTTCCATAGTATTCAGTTGACATGATCAAGTATTACTCTTTCTTTT 60  
DB 123720 CTCTCTATTTTCCATAGTATTCAGTTGACATGATCAAGTATTACTCTTTCTTTT 123779  
QY 61 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 120  
DB 123780 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 123839  
QY 121 TGGCATGATAGTCTGATGCACTCCACCTCTCTGAGTTCAAGCAAAAGCTGCTCCCTC 180  
DB 123840 TGGCATGATAGTCTGATGCACTCCACCTCTCTGAGTTCAAGCAAAAGCTGCTCCCTC 123899  
QY 181 AGCTCCCGGGTATGATGGAATTACAGGCGCCACACACACACACCTGGCTAATGTTGTAAT 240  
DB 123900 AGCTCCCGGGTATGATGGAATTACAGGCGCCACACACACACACCTGGCTAATGTTGTAAT 123959  
QY 241 TTTAGTAGAGATGGGGTTTTCACCATGTTGGCCAGGCTGCTCAAACTCTGACCTCAGA 300  
DB 123960 TTTAGTAGAGATGGGGTTTTCACCATGTTGGCCAGGCTGCTCAAACTCTGACCTCAGA 124019  
QY 301 G 301  
DB 124020 G 124020

RESULT 6  
US-60-465-241-52403  
Sequence 52403, Application US/60465241  
GENERAL INFORMATION:  
APPLICANT: CARGILL, Michele  
APPLICANT: BEGOVICH, Ann  
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
TITLE OF INVENTION: RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: C1001468  
CURRENT APPLICATION NUMBER: US/60/465,241  
CURRENT FILING DATE: 2003-04-23  
NUMBER OF SEQ ID NOS: 258418  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 52403  
LENGTH: 22683  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-60-465-241-52403

Query Match 99.6%; Score 299.8; DB 103; Length 22683;  
Best Local Similarity 99.0%; Pred. No. 1.6e-25;  
Matches 298; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCTCTATTTTCCATAGTATTCAGTTGACATGATCAAGTATTACTCTTTCTTTT 60  
DB 10179 CTCTCTATTTTCCATAGTATTCAGTTGACATGATCAAGTATTACTCTTTCTTTT 10238  
QY 61 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 120  
DB 10238 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 10298  
QY 121 TGGCATGATAGTCTGATGCACTCCACCTCTCTGAGTTCAAGCAAAAGCTGCTCCCTC 180  
DB 10299 TGGCATGATAGTCTGATGCACTCCACCTCTCTGAGTTCAAGCAAAAGCTGCTCCCTC 10358

QY 181 AGCTCCCGGGTATGATGGAATTACAGGCGCCACACACACACACCTGGCTAATGTTGTAAT 240  
DB 10359 AGCTCCCGGGTATGATGGAATTACAGGCGCCACACACACACACCTGGCTAATGTTGTAAT 10418  
QY 241 TTTAGTAGAGATGGGGTTTTCACCATGTTGGCCAGGCTGCTCAAACTCTGACCTCAGA 300  
DB 10419 TTTAGTAGAGATGGGGTTTTCACCATGTTGGCCAGGCTGCTCAAACTCTGACCTCAGA 10478  
QY 301 G 301  
DB 10479 G 10479

RESULT 7  
US-60-470-166-8933  
Sequence 8933, Application US/60470166  
GENERAL INFORMATION:  
APPLICANT: CARGILL, Michele  
APPLICANT: GRUBE, Andrew  
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH  
TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF  
FILE REFERENCE: C1001464  
CURRENT APPLICATION NUMBER: US/60/470,166  
CURRENT FILING DATE: 2003-05-09  
NUMBER OF SEQ ID NOS: 68617  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 8933  
LENGTH: 22683  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-60-470-166-8933

Query Match 99.6%; Score 299.8; DB 104; Length 22683;  
Best Local Similarity 99.0%; Pred. No. 1.6e-25;  
Matches 298; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCTCTATTTTCCATAGTATTCAGTTGACATGATCAAGTATTACTCTTTCTTTT 60  
DB 10179 CTCTCTATTTTCCATAGTATTCAGTTGACATGATCAAGTATTACTCTTTCTTTT 10238  
QY 61 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 120  
DB 10238 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 10298  
QY 121 TGGCATGATAGTCTGATGCACTCCACCTCTCTGAGTTCAAGCAAAAGCTGCTCCCTC 180  
DB 10299 TGGCATGATAGTCTGATGCACTCCACCTCTCTGAGTTCAAGCAAAAGCTGCTCCCTC 10358  
QY 181 AGCTCCCGGGTATGATGGAATTACAGGCGCCACACACACACCTGGCTAATGTTGTAAT 240  
DB 10359 AGCTCCCGGGTATGATGGAATTACAGGCGCCACACACACACCTGGCTAATGTTGTAAT 10418  
QY 241 TTTAGTAGAGATGGGGTTTTCACCATGTTGGCCAGGCTGCTCAAACTCTGACCTCAGA 300  
DB 10419 TTTAGTAGAGATGGGGTTTTCACCATGTTGGCCAGGCTGCTCAAACTCTGACCTCAGA 10478  
QY 301 G 301  
DB 10479 G 10479

RESULT 8  
US-09-947-911-286  
Sequence 286, Application US/09947911  
GENERAL INFORMATION:  
APPLICANT: VENTER, J. Craig  
TITLE OF INVENTION: COLLECTION OF SINGLE NUCLEOTIDE POLYMORPHISM (SNPs) LOCATED ON  
FILE REFERENCE: C1001300  
CURRENT APPLICATION NUMBER: US/09/947,911  
CURRENT FILING DATE: 2001-09-07  
NUMBER OF SEQ ID NOS: 368



```

1 SEQ ID NO 286
2
3 LENGTH: 3037227
4
5 TYPE: DNA
6
7 ORGANISM: HUMAN
8
9 FEATURE:
10
11 NAME/KEY: misc_feature
12
13 LOCATION: (1)...(3037227)
14
15 OTHER INFORMATION: n = A,T,C or G
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
102
```

Query Match	69.4%;	Score 208.8;	DB 39;	length 3037227;
Best Local Similarity	88.6%;	Pred. No. 1.1e-15;		
Matches 225; Conservative	1;	Mismatches 28;	Indels 0;	Gaps 0;

Oy	4	TTCACCTTCCCTTTTCTTTTTTTTTTTTTTTTGAGATGCAATTGGTCTGTGACC	105
Dd	1512582	TTCTCTTCCCTTTTTTTTTTTTTTTTGAGATGCAATTGCCTCTGTGCC	1512641

Qy	10	CATGTGGAGTGAATGGCATGATCATACTCACTGCAACTCCACCTCTGGATTCAAG	165
Db	1512642	CAGCTGGAGTGCATAGGCTGATCTCACTCACTGCAACTCCACCTCTCTGGATTCAAG	1512700

Db	1512702	CGATTCTCCGCGCTCAGCCTCTTCAAGTAGTGGATTACAGGTCCTTCCGCCACCAACCCA	1512761
Oy	166	CAAAGCTGTGCGCTCAGCCTCCCGGGTAAATGGATTACAGGCGGCCACACACACACTCG	225

Accession	Gene	Length (bp)
OY 226	GCTATGTTTGTATTTTGTAGAGATGGGGTTTCACACATGTTGGCCACGGCTGGTCTCAA	285
Db 1512762	GCTAATCTTTGTATTTTGTAGAGACAGGGTTTCACCATGTTGGCCACGGCTGGTCTCGA	1512822

[illegible]

RESULT 9  
PCF-US01-21148-3/C  
Sequence 3 Amplification PC/TMS0121148

```

; GENERAL INFORMATION: Gennady et al.
; APPLICANT: MERKULOV,
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEININs

```

```

1      TITLE OF INVENTION:  AND USES THEREOF
2
3      FILE REFERENCE:  CL000723PCT
4
5      CURRENT APPLICATION NUMBER:  PCT/US01/21148
6
7      CURRENT FILING DATE:  2001-07-03

```

PRIOR APPLICATION NUMBER: 60/216,340  
 PRIOR FILING DATE: 2000-07-05  
 PRIOR APPLICATION NUMBER: 09/810,673  
 PRIOR FILING DATE: 2001-03-19

```

: NUMBER OF SEQ ID NOS: 7
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 3
: LENGTH: 26657

```

```

!
! IFE: DNA
! ORGANISM: Human
PCT-US01-21148-3

```

Query Match	87.68;	Denglin 2005;
Best Local Similarity	87.48;	
Matches 222; Conservative	1; Mismatches	0; Gaps

Db 445 TTTTCTTTTTTTTTTTTTTTTTTTTGAGATGAGGTTCGCTCTATTG 386

385 CCGAGGCTGGAGTGCATGGCAGCATCTCAGCTCACCACTCCGCTCCCGGGATTCA 326

325 AGCAATGCTCCCGCCTCAGCCTCCCAAGTAGCTGAGATTACAGGCGCACACCAACGCC 286

Accession	Sequence	Length
Oy	CGGCTATGTTTATTTTATGAGAGATGGGGTTTCCATGTTGGCCAGGCTGCTC	287
Db	CGGCTATTTTGTATTTTATTTAGTAGAGACGGGGTTCCATGTTGGTTAGGCTGCTC	206
Oy	22 CGGCTATGTTTATTTTATGAGAGATGGGGTTTCCATGTTGGCCAGGCTGCTC	287
Db	265 CGGCTATTTTGTATTTTATTTAGTAGAGACGGGGTTCCATGTTGGTTAGGCTGCTC	206
Oy	284 AAATCTCTGACCTC	297
Db	205 AAATCTCTGACCTC	192

RESULT 10  
US-09-810-673A-3/c  
; Sequence 3, Application US/09810673A  
ORIGINAL TRANSMISSION.

APPLICANT: MERKULOV, Gennady et al.  
TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,  
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS  
TITLE OF INVENTION: AND USES THEREOF

```

; FILE REFERENCE: CL000723
; CURRENT APPLICATION NUMBER: US/09/810,673A
; CURRENT FILING DATE: 2001-03-19
; NUMBER OF SEQ ID NOS: 7

```

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 26657
; TYPE: DNA

```

Query Match	67.8%	Score 204	DB 33	Length 26657
!	ORGANISM: Human			
US-09-810-673A-3				

44	TATACCTCTTCTTTCTTTTCTTTTGAGATGAGATTGGCTCTGTG	103
Best Local Similarity	87.4%;	Pred. NO. 1e-14;
Matches	222;	Conservative 1; Mismatches 31; Indels 0; Gaps 0

Db 445 TTTTTTTTTTTTTTTTTTTTTTTTGAGATGAGTTCCGCTTATTG 38

Db 385 CCCAGGCTGAGTGCATGCGACGATCTCAGCTCACACACACTCCGCTCCGGGTCA 322

Db 325 AGCATGCTCCCGCCTCAGCCTCCCAAGTAGCTGAGATTACAGCGCACACCAACGCC 26

Oy 224 CGGCTAATGTTTGATTTTATAGTAGATGGGGTTTCCACCATGTTGGCCAGGCTGGTCTC 28

Db 265 CGGCTAATTTTGTATTTTGTAGACGCGGGTTTCAACCAATGTTGGTTAGGCTGCTC 20

Qy 284 AAACCTCCTGACCTC 297

Db 205 AAACCTCCTGACCTC 192

US-10-395-242-3/c  
; Sequence 3, Application US/10395242  
; GENERAL INFORMATION:

TITLE OF INVENTION: ISOLATED HUMAN ION CHANNEL PROTEINS,  
 TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN ION CHANNEL PROTEINS,  
 TITLE OF INVENTION: AND USES THEREOF

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```

; SEQ ID NO 3
;
; LENGTH: 26657
; TYPE: DNA

```

US-10-395-242-3



	Query March	67.8%;	Score 204;	DB 77;	Length 32768;
	Best Local Similarity	87.4%;	Pred. No. 9.7e-15;		
	Matches 222;	Conservative 1;	Mismatches 31;	Indels 0;	Gaps 0;
QY	44 TATTACTCTTTCCTTTTCTTTTTTTTTTTTTTTTTTTTTTTTGAGATGACGTTTGGCTTTGTG	103			
Dd	26563 TTGAGATGAGATTTCCGCTTAATG	26622			
QY	104 CCCATGCTGAGTGAATGCATGAYCANGCTCACTGCAACTCCACTCCTGGATTCA	163			
Dd	26623 CCCAGGCTGGAGTGCATAGGACGATCTCAGTCAACAACCTCGCCTCCGGGTTCA	26882			
QY	164 AGCAAAGCTGTCGCTTAGCCTCCCGGGTNGATGGGATTTACAGCGGCCAACCAACCACT	223			
Dd	26683 AGCATAGCTCCGCGCTCAGCTCCCAAGTAGCTGAGATTTACAGGCGCACACCAACCAACGCGC	26742			
QY	224 CGGCTAATGTTGATTTTAACTAGAGAAATGGGGTTTACCATGTTGGCAGGCTGGTCTC	283			
Dd	26743 CGGCTAATTTTGTATTTTATTTTAAAGAGAAGCGGGTTTCCCATGTGTGGTTAGGCTGGTCTC	26802			
QY	284 AAACCTCCGACCTC 297				
Dd	26803 AAACCTCCGACCTC 26816				

```

      RESULT 14
US-09-947-916-16/c
; Sequence 16, Application US/09947916
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: COLLECTION OF SINGLE NUCLEOTIDE POLYMORPHISM (SNPs) LOCATED ON
; TITLE OF INVENTION: 19, METHODS OF DETECTION, AND USES THEREOF
; FILE REFERENCE: CLO001296
; CURRENT APPLICATION NUMBER: US/09/947,916
; CURRENT FILING DATE: 2003-03-20
; NUMBER OF SEQ ID NOS: 337
; SEQ ID NO 16
; LENGTH: 7928029
; TYPE: DNA
; ORGANISM: HUMAN
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(7928029)
; OTHER INFORMATION: n = A,T,C or G
US-09-947-916-16

Query Match          67.8%; Score 204; DB 39; Length 7928029;
Best Local Similarity 87.4%; Pred. No. 3.1e-15;
Matches 222; Conservative 1; Mismatches 31; Indels 0; Gaps 0;

```



OY	44	TATTAACCTTCTTTTTTTTTTTTTTTTTTTTTTTTTGAGATGGAGTTTGTCGTGTG	103
Db	2278243	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGAAGTAGAGTTTGCCTCTATTG	2278164
OY	104	CCCATGCTGGAGTGGAAATGGCATAATGCTCACTGCACCTCTCTGGGTCCA	163
Db	2278183	CCCAGCGCTGGAGTGGAAATGGCACAATGCTCACTGCACACTCTCCGGGTCCA	2278124
OY	164	AGCAAAGCTGTGCGCTCAGACCCTCCGGGGTAATGGATTACAGGCCGCCACACACACT	223
Db	2278123	AGCAATGCTCCCGCTCAGCCTCCCAAGTAGCTGATTAAGAAGGCGCACACACACGCC	2278064
OY	224	CGGCTAATGTTTGTAATTTTGTAGTAGAGATGGGGTTTCCCATGTTGGCCAGGCTGGTCTC	283
Db	2278063	CGGCTAATTTTGTATTTTGTAGTAGAGACGGGGTTTCCCATGTTGGTTAGGCTGGTCTC	2278004
OY	284	AAACTCTGACCTC 297	
Db	2278003	AAACTCTGACCTC 2277990	

RESULT 15

```

US-60-206-600-43/C
Sequence 43, Application US/60206600
GENERAL INFORMATION:
APPLICANT: Beasley, Ellen
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS OF THE
TITLE OF INVENTION: SER/THR AND TYR FAMILY OF KINASES, NUCLEIC ACID MOLECULES
TITLE OF INVENTION: ENCODING THESE HUMAN KINASE PROTEINS, AND USES THEREOF
FILE REFERENCE: CL000627
CURRENT APPLICATION NUMBER: US/60/206,600
CURRENT FILING DATE: 2000-05-31
NUMBER OF SEQ. ID NOS: 172
SOFTWARE: PaatSeq for Windows Version 4.0
SEQ ID NO 43
LENGTH: 32768
TYPE: DNA
ORGANISM: HUMAN
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)..(32768)
OTHER INFORMATION: n = A,T,C OR G
US-60-206-600-43

```

Query Match	67.4%;	Score 202.8;	DB 76;	Length 32768;
Best Local Similarity	85.2%;	Pred. No. 1.3e-14;		
Matches 225;	Conservative 1;	Mismatches 38;	Indels 0;	Gaps 0;

[illegible]

```
Search completed: March 12, 2004, 09:13:14
Job time : 2466.15 secs
```



1	198.8	66.0	393141	6	US-10-767-471-10617	Sequence 10617, A
2	197.2	65.5	3878	6	US-10-100-663-11260	Sequence 11260, A
3	193	66.1	154875	6	US-10-765-790-107	Sequence 107, A
4	192.4	63.9	30449	6	US-10-767-471-10902	Sequence 10902, A
5	192	63.8	58905	7	US-60-548-091-5611	Sequence 5611, A
6	192	63.8	98311	6	US-10-767-471-10822	Sequence 10822, A
7	191.6	63.7	546025	6	US-10-767-471-10717	Sequence 10717, A
8	191	63.5	401433	6	US-10-765-790-79	Sequence 79, A
9	190.8	63.4	71919	6	US-10-767-471-10899	Sequence 10899, A
10	190	63.1	24974	6	US-10-767-471-10720	Sequence 10720, A
11	190	63.1	63124	6	US-10-417-315A-82	Sequence 82, A
12	189.4	62.9	13885	6	US-10-767-471-10735	Sequence 10735, A
13	189.2	62.9	1989817	6	US-10-767-471-10675	Sequence 10675, A
14	189	62.8	12785	6	US-10-767-471-10711	Sequence 10711, A
15	189	62.8	36106	6	US-10-767-471-10713	Sequence 10713, A
16	189	62.8	144021	6	US-10-767-471-10552	Sequence 10592, A
17	188.8	62.7	1789817	7	US-60-548-091-6641	Sequence 5641, A
18	188.8	62.7	228854	7	US-60-548-091-6633	Sequence 5633, A
19	188.6	62.7	107507	6	US-10-767-471-10579	Sequence 10579, A
20	188.4	62.6	44594	6	US-10-767-471-10875	Sequence 10875, A
21	188.4	62.6	344805	6	US-10-767-471-1071	Sequence 1, A
22	188.4	62.6	354592	6	US-10-765-790-70	Sequence 70, A
23	188.2	62.5	149046	7	US-60-548-091-6651	Sequence 5651, A
24	188	62.5	803755	6	US-10-767-471-10700	Sequence 10700, A
25	187.6	62.3	103375	6	US-10-417-315A-94	Sequence 94, A
26	187.4	62.3	448234	6	US-10-767-471-10706	Sequence 10706, A

C	27	187.2	62.2	391.05	7	US-60-5548-091-5534	Sequence 5634, App
C	28	187.2	62.2	616.55	7	US-10-765-790-1105	Sequence 1055, App
C	29	187.2	62.2	616.55	7	US-10-765-790-1118	Sequence 118, App
C	30	187.2	62.2	750.07	7	US-60-5448-091-5501	Sequence 5601, App
C	31	187	62.1	1004.8	6	US-10-100-683-11509	Sequence 11509, App
C	32	187	62.1	3448.05	6	US-10-779-271-1	Sequence 1, Appl
C	33	187	62.1	3545.92	6	US-10-765-790-70	Sequence 70, Appl
C	34	187	62.1	1790.242	6	US-10-767-471-10805	Sequence 10805, Appl
C	35	186.8	62.1	646.69	6	US-10-767-471-10845	Sequence 10845, Appl
C	36	186.2	61.9	188.53	6	US-10-767-471-10737	Sequence 10737, Appl
C	37	186.2	61.9	720.69	6	US-10-775-169-82	Sequence 82, Appl
C	38	186	61.8	1279.17	6	US-10-767-471-10827	Sequence 10827, Appl
C	39	186	61.8	2347.34	7	US-60-548-091-5703	Sequence 5703, App
C	40	185.8	61.7	2301.01	6	US-10-767-471-10691	Sequence 10691, App
C	41	185.8	61.7	1230.54	6	US-10-767-471-10655	Sequence 10655, App
C	42	185.6	61.6	1660.20	6	US-10-767-471-10806	Sequence 10806, App
C	43	185.4	61.6	1067.07	1	PCT-US93-44223-8	Sequence 8, Appl
C	44	185.2	61.5	1483.86	6	US-10-767-471-10578	Sequence 10578, App
C	45	185.2	61.5	1483.86	6	US-10-767-471-10578	Sequence 10578, App

## ALIGNMENTS

## RESULT 1

```

Sequence 1067, Application US/10767471
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele et al.
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
TITLE OF INVENTION: RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CL001505
CURRENT APPLICATION NUMBER: US/10/767,471
CURRENT FILING DATE: 2004-01-30
NUMBER OF SEQ. ID NOS: 50231
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 10617
LENGTH: 393141
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(393141)
OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tabl
IS-10-767-471-10617

```

## Query Match

Best Local Similarity 85.8%; Pred. No. 9.6e-30;  
Matches 218; Conservative 2; Mismatches 34; Indels 0; Gaps 0.

[illegible]



















